# TAILS N-terminomics and proteomics reveal complex regulation of proteolytic cleavage by *O*-glycosylation

Received for publication, January 18, 2018, and in revised form, March 27, 2018 Published, Papers in Press, March 28, 2018, DOI 10.1074/jbc.RA118.001978

**Sarah L. King[‡1], Christoffer K. Goth[‡1], Ulrich Eckhard[§], Hiren J. Joshi[‡], Amalie D. Haue[‡], Sergey Y. Vakhrushev[‡], Katrine T. Schjoldager[‡], Christopher M. Overall[§2], and Hans H. Wandall[‡3]**

*From the [‡]Department of Cellular and Molecular Medicine, Centre for Glycomics, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen, Denmark and the [§]Centre for Blood Research, Department of Oral Biological and Medical Sciences, Faculty of Dentistry, and Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada*

Edited by Gerald W. Hart

**Proteolytic processing is an irreversible post-translational modification functioning as a ubiquitous regulator of cellular activity. Protease activity is tightly regulated via control of gene expression, enzyme and substrate compartmentalization, zymogen activation, enzyme inactivation, and substrate availability. Emerging evidence suggests that proteolysis can also be regulated by substrate glycosylation and that glycosylation of individual sites on a substrate can decrease or, in rare cases, increase its sensitivity to proteolysis. Here, we investigated the relationship between site-specific, mucin-type (or GalNAc-type) *O*-glycosylation and proteolytic cleavage of extracellular proteins. Using *in silico* analysis, we found that *O*-glycosylation and cleavage sites are significantly associated with each other. We then used a positional proteomic strategy, terminal amine isotopic labeling of substrates (TAILS), to map the *in vivo* cleavage sites in HepG2 SimpleCells with and without one of the key initiating GalNAc transferases, GalNAc-T2, and after treatment with exogenous matrix metalloproteinase 9 (MMP9) or neutrophil elastase. Surprisingly, we found that *loss* of GalNAc-T2 not only increased cleavage, but also decreased cleavage across a broad range of other substrates, including key regulators of the protease network. We also found altered processing of several central regulators of lipid homeostasis, including apolipoprotein B and the phospholipid transfer protein, providing new clues to the previously reported link between *GALNT2* and lipid homeostasis. In summary, we show that loss of GalNAc-T2 *O*-glycosylation leads to a general decrease in cleavage and that GalNAc-T2 *O*-glycosylation affects key regulators of the cellular proteolytic network, including multiple members of the serpin family.**

Proteases are tightly regulated hydrolytic enzymes that act to irreversibly and specifically cleave peptide bonds. In humans, over 560 proteases (1) exist that can be grouped into serine, cysteine, aspartic, metallo, and threonine endopeptidases based on their catalytic mechanism (2). Endopeptidases constitute the second largest enzyme class and are the target of ~5–10% of drug therapies (3–5). Proteases control activity, turnover, and localization of proteins, many of which are key signaling molecules or cytokines, and contribute to host defense or tissue invasion and exotoxin activity in infection, immunity, homeostasis, differentiation, migration, and cell cycle progression (5, 6). Proteases are of particular importance in the extracellular space. Here, limited and specific proteolysis, but not complete degradation of targets, maintains a dynamic environment in response to changing conditions (5). Such cleavages activate or inactivate proteins (7), release receptors from the membrane surface (8), and alter or even invert protein function (9–11).

Understanding the specific targets of a protease is crucial for determining its function. However, defining physiologically relevant protein substrates and cleavage sites for each protease is nontrivial. Protease substrates can be identified *in vitro* using biochemical assays and high-throughput screens, whereas yeast two-hybrid methods or immobilized inactive protease domains can be used to probe more complex exosite-driven interactions (5, 12). Using these methods, homologous proteases often display broad and overlapping specificities, which are not necessarily recapitulated *in vivo*. This disparity arises because, in the cellular setting, protease activity is tightly regulated by control of gene expression; compartmentalization and sequestering of enzyme activity; zymogen activation and enzyme inactivation; substrate availability and specificity; and post-translational

This work was supported in part by Danish Research Councils Grant 1331-00133B (to S. L. K. and H. H. W.); by a Programme of Excellence 2016 (Copenhagen as the Next Leader in Precise Genetic Engineering CDO2016: 2016CDO04210) from the University of Copenhagen (to S. L. K., C. K. G., A. D. H., and H. H. W.); by Danish National Research Foundation Grant DNRF107 (to S. L. K., C. K. G., A. D. H., H. J. J., S. Y. V., K. T. S., and H. H. W.); and by the NEYE Foundation (to C. K. G.). This work was also supported by Canadian Institutes of Health Research Grant MOP-37937 (to C. M. O.) and by the British Columbia Proteomics Network (to C. M. O.). The authors declare that they have no conflicts of interest with the contents of this article.

*The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (77) with the data set identifier PXD009453.*

This article contains Figs. S1 and S2.

[1] Both authors contributed equally to this work.

[2] Holder of a Canada Research Chair in Protease Proteomics and Systems Biology. To whom correspondence may be addressed: Centre for Blood Research, Faculty of Dentistry, and Dept. of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. Tel.: 604-822-3561; E-mail: chris.overall@ubc.ca.

[3] To whom correspondence may be addressed: Dept. of Cellular and Molecular Medicine, Centre for Glycomics, University of Copenhagen, DK-2200 Copenhagen N, Denmark. Tel.: 45-2721-0936; E-mail: hhw@sund.ku.dk.

modification (PTM)[4] of the enzyme or substrate. For example, lysine methylation is reported to protect substrates from cleavage by the widely used endoproteinase Lys-C (13). Another example is the positive and negative regulation of cleavage by caspase-3, -7, and -8 during apoptosis by substrate phosphorylation (14, 15).

Substrate glycosylation is also suggested to alter substrate sensitivity to proteolysis. *N*-Glycans on subtilisin and matriptase have been shown to protect these proteins from proteolytic cleavage (16–18). Similarly, dense mucin-type or *O*-*N*-acetylgalactosamine (*O*-GalNAc) *O*-glycosylation (herein referred to as *O*-glycosylation) has long been proposed to nonspecifically shield proteins from degradation in the extracellular space (19). However, it is increasingly apparent that individual *O*-glycosites present on substrates can also exert fine control over protease specificity. The first example of this was the inhibition of proprotein processing of pro-opiomelanocortin by *O*-glycosylation at Thr-45 (20–22). More recently, increased proprotein processing resulting from loss of *O*-glycosylation has been associated with familial tumoral calcinosis (23) and serum dyslipidemia (24, 25). Such cross-talk between *O*-glycosylation and proprotein convertases may be a general phenomenon, as peptide assays indicate that *O*-glycans can inhibit cleavage by furin on multiple substrates (26). Furthermore, regulation of proteolysis by *O*-glycosylation is not limited to an individual protease family, as shedding by members of the A disintegrin and metalloproteinase (ADAM) subfamily is also inhibited by site-specific *O*-glycosylation (27, 28). Indeed, *O*-glycosylation of a growing number of substrates has been shown to alter their sensitivity to proteases of the aspartic ($\beta$-secretase), serine (furin, chymotrypsin), or metallo (ADAMs and metalloproteinases (MMPs)) families, suggesting that cross-talk between these PTMs may occur independently of the mechanism of catalysis (29). However, this does not indicate that *O*-glycosylation is an indiscriminate inhibitor of proteolysis. Indeed, glycosylation of specific Ser or Thr residues may generate disparate responses ranging from complete inhibition to activation even within the same protease family, as has been demonstrated in the case of ADAM proteases (28). Moreover, *in vitro* peptide assays and cell-based reporter constructs support a model in which the *O*-glycan is located within or close to the binding cleft, and, at least in the case of furin, the closer the glycan is to the scissile bond, the greater its inhibitory effect (26). Given that recent developments in MS, within our laboratory and others, have shown that over 85% of proteins transiting through the secretory pathway are likely to be *O*-glycosylated (30), cross-talk between these PTMs could have widespread implications.

In the present study, we investigated the relationship between site-specific *O*-glycosylation and proteolytic cleavage of extracellular proteins. *O*-Glycosylation is unique, as it is initiated by a family of 20 *N*-acetylgalactosamine transferases (GalNAc-Ts), each of which displays a partly overlapping but distinct substrate specificity (31–33). Although GalNAc-T enzymes have not been shown to respond dynamically to changes in the cellular environment to date, they display markedly different expression profiles between cell types and during differentiation (34–36). Thus, the same protein expressed by two different cells can be identical in sequence but differ in its *O*-glycosylation (37–39). Given the emerging evidence that glycosylation can alter substrate sensitivity to proteolysis, we hypothesized that alternate site-specific glycosylation could be used to globally fine-tune protein sensitivity to proteolysis in a protease- and cell-specific manner. To investigate whether there was evidence for this hypothesis, we first performed a comprehensive bioinformatic analysis of known *O*-glycosylation sites with known proteolytic cleavage sites in the entire human proteome and found a significant association between *O*-glycosylation sites and cleavage sites. We next asked whether lack of a subset of *O*-glycosylation sites would alter the global proteolytic cleavage pattern, and if so, whether we could identify certain classes of proteolytic cleavage sites that were particularly affected. To address this question, we used a robust positional proteomic strategy, terminal amine isotopic labeling of substrates (TAILS) (40, 41) to identify cleavage sites in cells with and without GalNAc-T2 expression. GalNAc-T2 has been shown to be a major driver of *O*-glycosylation (42), and several examples of interplay between GalNAc-T2–specific *O*-glycosylation and proteolytic cleavage have been described (27, 28, 43), making GalNAc-T2 a prime target for this approach. With this strategy, we identified 189 proteolytic cleavage sites altered by the lack of a subset of *O*-glycosylation sites. We identified substrates where loss of *O*-glycans caused increased cleavage, indicating an expected shielding of the cleavage site, and, most unexpectedly, we also identified a second subset of substrates where loss of *O*-glycans caused decreased cleavage. Overall, our studies reveal that *O*-glycosylation can both decrease and also increase substrate cleavage.

## Results

### *In silico identification of cleavage sites proximal to O-glycosylation sites*

Despite evidence of cross-talk between GalNAc-type *O*-glycosylation and mainly proprotein convertases and ADAM metalloproteases, it is unknown which other proteases are subject to regulation by site specific *O*-glycosylation. Moreover, it is not clear to what degree site-specific *O*-glycosylation contributes to the overall substrate degradome of a particular protease. To address these questions, we first identified candidate proteases that targeted *O*-glycoproteins. We performed an unbiased *in-silico* screen using the TopFIND knowledgebase (44), the protein termini and cleavage site knowledge base, and selected proteases with well-defined degradomes (>40 cleavages in >20 proteins). We found that all proteases cleaved *O*-glycoproteins. However, the proportion of *O*-glycoprotein substrates varied from 2 to 69% (Fig. 1*A*). MMPs, neutrophil elastase, thrombin, plasmin, kallikrein, and proprotein convertases were among the proteases that targeted *O*-glycoproteins most frequently, with *O*-glycoproteins constituting >35% of their substrates.

[4] The abbreviations used are: PTM, post-translational modification; ADAM, A disintegrin and metalloproteinase; MMP, matrix metalloproteinase; GalNAc-T, GalNAc transferase; TAILS, terminal amine isotopic labeling of substrates; SC, SimpleCells; FDR, false discovery rate; TIMP, tissue inhibitor of metalloproteinases; AMPN, aminopeptidase N; GO, gene ontology; KO, knockout; HPG-ALDII, hyperbranched polyglycerol polyaldehyde-derivatized polymer.

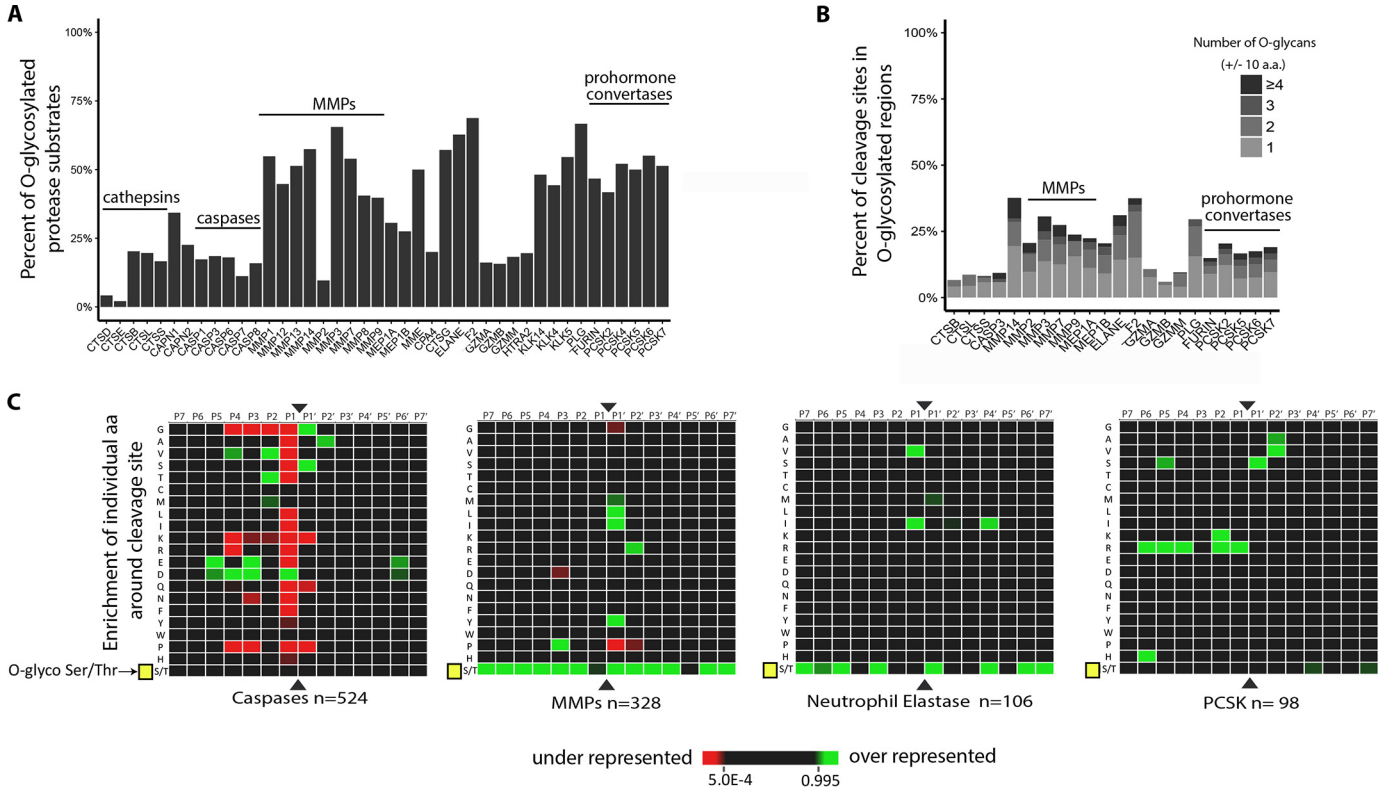


**Figure 1. *In silico* identification of the intersection of proteolysis and *O*-glycosylation.** Well-characterized proteases with >40 cleavage sites in >20 proteins were selected using the TopFIND database (69), and *O*-glycosylation of the substrates was interrogated using *O*-glycosite data from the glycodomain database (68). Proteases are listed by MEROPS family and official gene symbol. *A*, *O*-glycosylation of protease substrates shown per protease, indicating that a large proportion of MMP, serine protease, and prohormone convertase (*PCSK*) substrates are actually *O*-glycoproteins. *B*, occurrence of *O*-glycosylation in the region ± 10 amino acids (*aa*) around the scissile bond (P1 ↓ P1′) on *O*-glycoprotein substrates. *Shading* represents the total number of *O*-glycans present across the interval, indicating the extent of *O*-glycosylation around cleavage sites. *C*, analysis of the sequences flanking protease cleavage sites (P7–P7′) to identify over- and underenrichment of individual amino acids. All known *O*-glycosylated serine and threonine residues in the human proteome were recoded as *X* (indicated by a *yellow square*), essentially treating these modified residues as unique amino acids. The cleavage site consensus for each protease family was then determined using heat maps produced by IceLogo version 2.1 (71). To minimize sequence bias, cleavage sites listed in the TopFIND database were compared with a background of "psuedo-cleavage sites" generated by random sampling of the proteins listed in TopFIND. Statistically significant ($p < 0.001$) under- or overenrichment of each site is represented by *red* or *green shading*, respectively.

Having identified a set of well-characterized proteases that were demonstrated to cleave *O*-glycoproteins, we next determined whether *O*-glycosylation occurred in the vicinity of the cleavage site. We found that on *O*-glycoprotein substrates, more than 25% of MMP (MMP3, -7, -9, and -14), neutrophil elastase, thrombin, and plasmin cleavage sites were flanked by *O*-glycans within 10 residues up- or downstream of the scissile bond. In many cases, flanking regions carried multiple *O*-glycosites (Fig. 1*B*). There was little evidence for prime or nonprime (N- or C-terminal) bias in *O*-glycosylation, and if multiple *O*-glycans were present, they were frequently located on both sides of the scissile bond. Given the abundance of *O*-glycosylation around this subset of cleavage sites, we next asked whether *O*-glycans were under- or overrepresented at specific positions. To achieve this, we conceptualized *O*-glycosylated amino acids as a unique amino acid, *X*, and assessed the frequency with which they occurred using heat maps (Fig. 1*C*). Analysis was performed both with and without correction for protein disorder, as *O*-glycosylation and proteolytic cleavage occur more often in disordered loops (30, 45). Similar results were obtained with both analyses; for simplicity, only disorder-corrected data are presented here. *O*-Glycosylation was overrepresented ($p <$ 0.001) near MMP, neutrophil elastase, plasminogen, and thrombin cleavage sites. Interestingly, this overenrichment occurred at the majority of positions across the consensus sequence. Importantly, we did not find enrichment of *O*-glycosylation around caspase or proprotein convertase cleavage sites. In the case of caspases, this is expected, as this family of proteases are exclusively cytosolic. However, as shown in Fig. 1 (*A* and *B*), proprotein convertase substrates are still *O*-glycosylated in up to 50% of the cases, with up to 25% in close proximity to cleavage sites, suggesting general functional cross-talk, as has been proposed previously (26). Together, these results indicate that *O*-glycosylation occurs more frequently than expected around selected MMPs and selected serine protease cleavage sites.

### *Enrichment of N termini using TAILS*

Although *O*-glycosites are enriched around a subset of extracellular protease cleavage sites, the functional effect of such co-localization cannot be predicted *in silico*. Therefore, to test the consequences of the observed association, we employed the well-validated TAILS (40, 41) positional proteomics approach to identify the N-terminome in cells with and without GalNAc-T2 expression (24) (Fig. 2). To simplify the mass spectrometric analysis of the processed peptides, we used our pre-

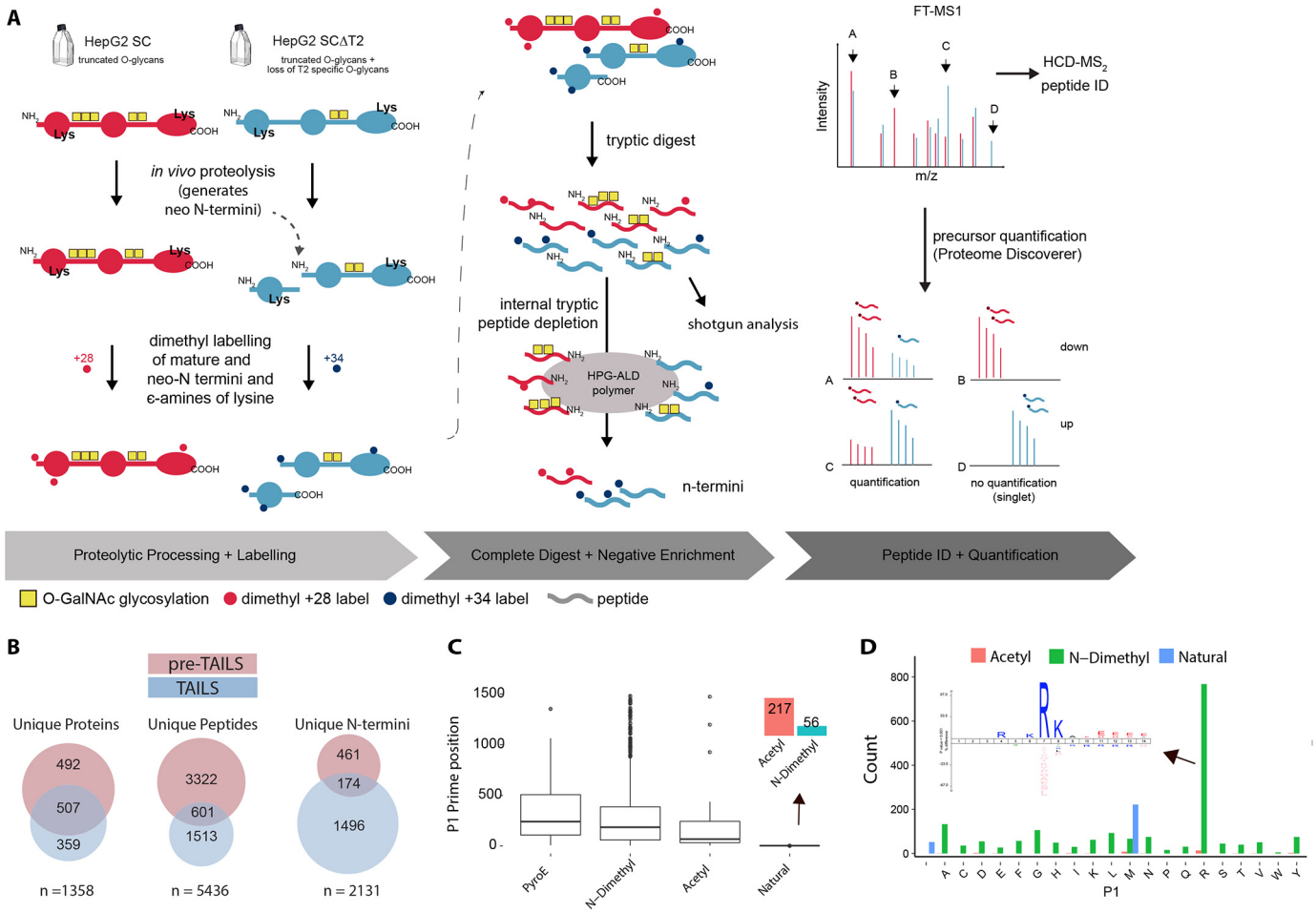


**Figure 2. Glyco-TAILS workflow for analysis of the interaction between limited proteolysis and *O*-glycosylation.** *A*, glycoengineered HepG2 cells lacking the galactosyltransferase-specific molecular chaperone C1GALT1C1 (SC) and the *O*-glycosyltransferases GALNT2 (SCΔT2) were subjected to N-terminal enrichment using TAILS. Unfractionated proteins were harvested from conditioned supernatant and subjected to *in vitro* cleavage using exogenous proteases. The free N termini of partially digested proteins (indicated in the *left panel*) were dimethyl-labeled to simultaneously block both endogenous and exogenously added protease-specific cleavage sites and to differentially label each sample. Samples were then combined and prepared for pre-TAILS shotgun-like MS by complete digestion using trypsin. The exposed amine groups of N termini generated by the trypsin digestion were removed by covalently coupling to a high-molecular weight polyaldehyde polyglycerol polymer. This allowed for selection via negative enrichment of blocked N termini (*middle panel*). Peptides were subsequently identified and quantified using high-resolution Orbitrap LC-MS/MS-HCD, combined with Proteome Discoverer version 2.1, and resultant data were analyzed in R using a purpose-built Glyco-TAILS workflow. *B*, counts of unique quantified proteins, peptides, and N termini pre- and post-TAILS enrichment in HepG2 SC. *C*, P1 prime position (the N-terminal amino acid of an identified N-terminal peptide) calculated from peptides identified in TAILS and mapped to a unique UniProt accession. *Numbers* indicate total group count. The graph is truncated at 1,500 on the ordinate axis. Termini identified at P1 or P2 were designated "natural"; the majority of these natural N termini were found to be acetylated (*inset*). *D*, abundance of amino acids at the P1 site for natural, acetylated, and *N*-dimethyl peptides, including (*inset*) IceLogo plot for peptides containing P1-Arg ($p < 0.001$).

viously described "SimpleCells" (SC) to secure a homogeneous *O*-glycoproteome (46). In SimpleCells, the first step in elongation of *O*-glycans is prevented due to elimination of the core 1 β3-galactosyltransferase-specific molecular chaperone (*COSMC*; *C1GALT1C1*). The inactivation of *COSMC* results in truncation of *O*-glycans, creating cells expressing only a simple *O*-glycan, the "Tn" antigen, composed of a single GalNAc monosaccharide attached to Ser/Thr residues. Using this approach, we identified 5,436 unique peptides corresponding to 1,358 proteins. In total, 2,131 unique N termini were identified with an FDR < 1.0% (Fig. 2*B*). 94% of unique peptides identified in the TAILS sample had blocked N termini (dimethyl-labeled, acetylated, or pyroglutamate), indicating efficient TAILS enrichment; of these peptides, 1,981 (93%) could be mapped to a unique accession in the SwissProt database, allowing determination of their positional identity. Of all of the peptides identified with a blocked N terminus, 83% carried an N-terminal dimethyl label, indicating that these peptides represented natural protein N termini or neo-N termini generated by protease cleavage. 11% of N termini peptides carried an acetyl group, and 12% of peptides mapped to natural N termini (either the initiating Met or position 2 of the canonical sequence), of which 79% were acetylated. A small proportion of acetylated N termini were located further within the protein, probably representing alternate start sites or post-translational acetylation after proteolytic processing (47) (Fig. 2*C*). We did not consider further internal peptides that initiated with an Arg in P1′ and also did not exhibit a dimethylated N terminus, as these would have been generated after tryptic cleavage with ~4–6% carrying through after polymer pull-out (Fig. 2*A*). Assessment of the P1 position for all identified N termini illustrated further differences from previous reports, as the P1 position was dominated by Arg (Fig. 2*D*). It is important to note that these were not due to tryptic cleavage in sample processing, as the identified pep-

tides all had a dimethylated N-terminal α-amine. The consensus sequence of P1-Arg– containing peptides closely resembles the consensus sequence of proprotein convertases, probably indicating a high degree of endogenous serine protease activity (Fig. 2*D*, *inset*).

### Determination of O-glycan– dependent changes in MMP9 and neutrophil elastase substrates

Having validated our ability to identify the N-terminome, we next determined the quantitative changes in the N-terminome due to loss of GalNAc-T2 by performing TAILS on media from HepG2 SC and HepG2 SC without GalNAc-T2 (SCΔT2) (SC *versus* SCΔT2; $n = 3$). Based on our *in silico* screen described above, we chose to investigate MMP9 and neutrophil elastase, representative members of the metalloprotease and serine protease families targeting *O*-glycosylated substrates with well-described and large substrate repertoires. Before we could interrogate the interaction with GalNAc-T2–mediated glycosylation sites, we first identified specific cleavage products after treatment with neutrophil elastase and MMP9 protease. After protease treatment of the cell supernatant with exogenous proteases, we identified 123 putative neutrophil elastase and 68 putative MMP9 protease-specific peptides in this data set (Fig. 3*B*) (48, 49). Comparison of these N termini with those reported in the TopFIND database indicated that the majority of sites were novel. However, a small subset had been previously described. In these cases, there was good agreement between TopFIND and our study with 5 of 5 putative MMP9 sites and 3 of 6 putative neutrophil elastase sites previously reported as cleaved by MMP9 and neutrophil elastase, respectively (Fig. 3*C*). The agreement between these data sets is striking, given the disparate samples analyzed and the relatively small number of documented MMP9 and neutrophil elastase substrates. To determine whether there was evidence for functional interaction between neutrophil elastase or MMP9-specific cleavage and site-specific *O*-glycosylation, we next assessed whether the protease-specific N termini identified were affected by loss of GalNAc-T2 (Fig. 3*D* and Table 1). For this purpose, we calculated significant ratio cut-offs using the experimental variation of the internal tryptic peptides. From this analysis, we defined five quantitative categories of peptides that changed in SCΔT2: peptides specific to SCΔT2 cells (*I*); peptides significantly increased in SCΔT2 cells (*II*); peptides unchanged in SCΔT2 cells (*III*); peptides significantly decreased in SCΔT2 cells (*IV*); and peptides absent in SCΔT2 cells (*V*) (Fig. 3*D*). We found 20 of the 123 (16%) candidate neutrophil elastase-specific sites to be significantly altered in SCΔT2 cells (Table 1). Of the sites affected by loss of GalNAc-T2, 14 (70%) were found in known *O*-glycoproteins. Seven of the 20 sites (35%) were located within *O*-glycosylated regions, hinting at the possibility of a direct interaction between *O*-glycosylation and cleavage. To graphically depict the detected changes in cleavage across individual proteins, we constructed cleavage maps showing the localization and quantitation of individual peptides (Fig. 4*A*). These cleavage maps provide a method for visualization of the distance to known *O*-glycosites and illustrate the quantitative changes in pre-TAILS samples for each protein.

The substrates in which sites were affected by loss of GalNAc-T2 included glypican-3, reelin, α-2-HS-glycoprotein (fetuin-A), tissue inhibitor of metalloproteinases 1 (TIMP1), aminopeptidase N (AMPN), complement C3, and complement C4. Altered cleavage at the detected sites would be expected to have functional effects, such as protease shedding (AMPN), protein maturation (fetuin-A and reelin), or loss of activity (complement C3). There was also evidence that loss of GalNAc-T2 glycosylation affected neutrophil elastase–associated cleavages with no known glycosylation in the region, as observed in $\alpha_2$-macroglobulin, carboxypeptidase E, and inter-α-trypsin inhibitor heavy chain H3 (ITHI3). However, the majority of changes observed in the neutrophil elastase–treated samples were found when considering the P1-Arg– containing peptides, resulting from indirect effects, such as changes in the interconnected network of proteases known as the protease web (5) (Fig. 3*E*). For example, 78-kDa glucose-regulated protein (HSPA5) and mannosyl-oligosaccharide 1,2-α-mannosidase IA (MAN1A1) exhibited such effects. In contrast to neutrophil elastase, only 68 peptides (lacking a P1-Arg) were found to be specific to the MMP9 treatment group (Fig. 3*E*). Just three of the 68 MMP9 targets with glycosylation in close proximity to the cleavage site were found to be significantly changed with altered GalNAc-T2 *O*-glycosylation, suggesting a lack of major interaction between GalNAc-T2 and MMP9 in SCs (Fig. 3*E*). However, these results should be interpreted with caution, as our application of stringent quality cutoffs filtered out a large number of low-confidence MMP9 data points, including some potential *bona fide* targets. Peptide maps of proteins exhibiting potential direct, distal, and indirect changes in neutrophil elastase–specific cleavage are shown in Fig. 4.

### Changes in the cellular N-terminome with loss of GalNAc-T2

Having determined the cleavage pattern after MMP9 and neutrophil elastase treatment in cells with and without GalNAc-T2 dependent *O*-glycosylation, we next assessed the effect on endogenous protease activity. We therefore reanalyzed the data, removing protease treatment as a factor, thereby allowing determination of treatment-independent cleavage events. In total, 189 peptides were found to be significantly changed with deletion of GalNAc-T2. 40 (21%) of the changed peptides were found to reside in known *O*-glycosylated regions. Furthermore, 57% of significantly altered substrates were known *O*-glycoproteins, and substrates were significantly more likely to be *O*-glycosylated than proteins with unaltered N termini ($p < 2.2e-16$, odds ratio = 5.7, 95% confidence interval 5.0–6.6). Similar to neutrophil elastase-specific targets, the majority (173; 92%) of endogenous cleavage products were found to be decreased or absent in SCΔT2 cells, indicating that loss of GalNAc-T2 decreases cleavage of these targets. Complete data, including treatment-specific effects, are illustrated using cleavage maps in Fig. S1. By substrate winnowing to identify high-confidence targets, 62 cleavage sites in 50 substrates were identified as alternately processed in SCΔT2 cells (Table 2). The most marked change in these data sets was in the hepatocellular carcinoma marker, glypican-3 (50), which was entirely absent in SCΔT2 cells. The extracellular protein matrilin-3 also displayed substantial changes with loss of GalNAc-T2, as all three identi-

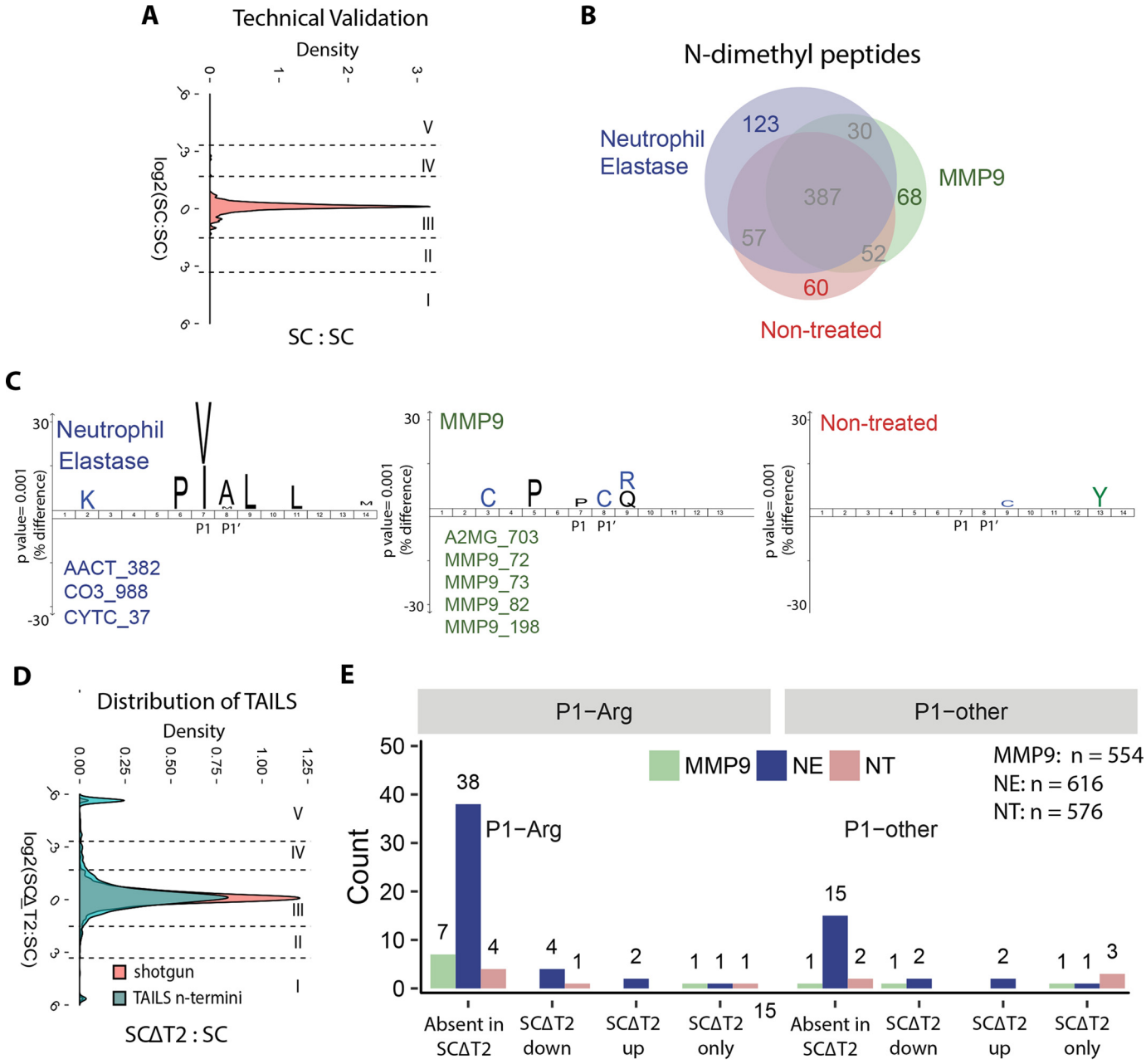


**Figure 3. Identification of treatment-specific cleavage events and effect of loss of GalNAc-T2.** *A*, quantitation and distribution of N termini from the control experiment comparing SimpleCells with SimpleCells (*SC:SC*), determining technical variation. The distribution of dimethyl ratios and the categories used to score the cleavage effect of loss of *O*-glycans in SimpleCells lacking GalNAc-T2 (SCΔT2 cells) in later experiments are shown. *B*, comparison of quantified N termini in each treatment group. Cleavage sites with a P1-Arg are excluded from the analysis to prevent loss of signal due to the extensive endogenous cleavage. The majority of N termini are identified in more than one treatment group. However, a subset of high-confidence peptides is unique to cleavage with either MMP9 or neutrophil elastase ("treatment-specific cleavage events"). *C*, IceLogo plots showing the consensus sequence of treatment-specific cleavage events, by comparing amino acid frequencies of treatment-dependent and treatment-independent N termini ($p$ value $< 0.001$). Previously reported MMP9 and neutrophil elastase–specific sites that are also identified in this study are listed *beside* the corresponding IceLogo plots. *D*, quantitation and distribution of N termini from SCΔT2:SC experiments (both with and without added proteases). *Dotted lines* indicate quantitative category thresholds as follows. *I*, peptides specific to SCΔT2 cells (KO only); *II*, peptides significantly increased in SCΔT2 cells (increased in KO); *III*, unchanged peptides (no change); *IV*, peptides significantly decreased in SCΔT2 cells (decreased in KO); *V*, peptides absent in SCΔT2 cells (absent in KO). Peptides identified in a single channel are arbitrarily assigned to the appropriate channel using an extreme quantitation ratio of 50:1 and therefore appear as separate peaks in *I* and *V*. *E*, count of significantly altered dimethyl-labeled N termini by treatment group. The total number of mapped peptides is indicated at the *top right*. Counts are separated into peptides with and without a P1-Arg.

fied N termini (PGR ↓ $^{51}$RPSPAAPD ↓ $^{59}$GAPA$^{63}$**S**G$^{65}$**T**$^{66}$**S** and VSR ↓ $^{109}$IID) were absent in SCΔT2 cells (where bold highlights possible glycosylation site). In both cases, N termini were found both distal and juxtaposed to *O*-glycosites. With the exception of matrilin-3 and glypican-3, all other proteins exhibited only localized changes with loss of GalNAc-T2 (when multiple cleavage sites were identified). For example, five cleavage regions were identified in angiotensinogen. However, GalNAc-T2–specific changes were only identified at KVL ↓ $^{182}$SAL ↓ $^{185}$QAV. These results suggest that loss of GalNAc-T2 does not alter total protein stability, but rather regulates limited proteolysis in a site-specific manner. In agree-

**Table 1**

**High-confidence neutrophil elastase– and MMP9 –specific sites significantly altered in SCΔT2 cells**

For each protein, the following information is provided in the respective columns: protease treatment, P1′ position, primary sequence surrounding the cleavage site, median quantification of T2 KO/WT ratio, number of times the peptide was quantified, and previously reported GalNAc-T2–specific glycosites. Residues in boldface red type are known *O*-glycosites, and red stretches are reported to be *O*-glycosylated, but without the exact site known. The underlining denotes the sequence of individual peptides identified.

| Entry name | UniProt | treatment | P1′ | P1 | | P1' | median quant | quant number |
|---|---|---|---|---|---|---|---|---|
| A2MG | P01023 | NE | 1431 | VSSNHVLIYLDKVSNQTLSLF | \| | FTVLQDVPVRDLKPAIVKVYDY | -5.64 | 2 |
| AMPN | P15144 | NE | 56 | QEKNKNANSSPVASTTPSASA | \| | TTNPASATTLDQSKAWNRYRLP | -5.64 | 2 |
| CBPE | P16870 | NE | 440 | YKLTASAPGYLAITKKVAVPY | \| | SPAAGVDFELESFSERKEEEKE | -5.64 | 4 |
| CO3 | P01024 | NE | 447 | TKKQELSEAEQATRTMQALPY | \| | STVGNSNNYLHLSVLRTELRPG | -5.64 | 3 |
| CO4A | P0C0L4 | NE | 1236 | NNLMAMAQETGDNLYWGSVTG | \| | SQSNAVSPTPAPRNPSDPMPQA | -5.64 | 2 |
| CO4A | P0C0L4 | NE | 1553 | SRECVGFEAVQEVPVGLVQPA | \| | SATLYDYYNPERRCSVFYGAPS | -2.6 | 2 |
| FETUA | P02765 | NE | 299 | PPSPPLGAPGLPPAGSPPDSH | \| | VLLAAPPGHQLHRAHYDLRHTF | -5.64 | 9 |
| GPC3 | P51654 | NE | 376 | RQYRSAYYPEDLFIDKKVLKV | \| | AHVEHEETLSSRRRELIQKLKS | -5.64 | 9 |
| RELN | P78509 | NE | 192 | DALAQQLCEQGAPTDVTVHPH | \| | LAEIHSDSIILRDDFDSYHQLQ | -5.64 | 2 |
| RELN | P78509 | NE | 3026 | YILLPEDALTNTTRLRWWQPF | \| | VISNGIVVSGVERAQWALDNIL | -5.64 | 2 |
| TIMP1 | P01033 | NE | 73 | VNQTTLYQRYEIKMTKMYKGF | \| | QALGDAADIRFVYTPAMESVCG | -5.64 | 3 |
| TRFE | P02787 | NE | 261 | DQYELLCLDNTRKPVDEYKDC | \| | HLAQVPSHTVVARSMGGKEDLI | -1.7 | 2 |
| CO3 | P01024 | NE | 1408 | EICTRYRGDQDATMSILDISM | \| | MTGFAPDTDDLKQLANGVDRYI | 5.64 | 2 |
| FETA | P02771 | MMP9 | 313 | QDTLSNKITECCKLTTLERGQ | \| | CIIHAENDEKPEGLSPNLNRFL | 5.64 | 2 |
| ITIH3 | Q06033 | NE | 560 | MEKALQERDYIFGNYIERLWA | \| | YLTIEQLLEKRKNAHGEEKENL | 1.9 | 5 |

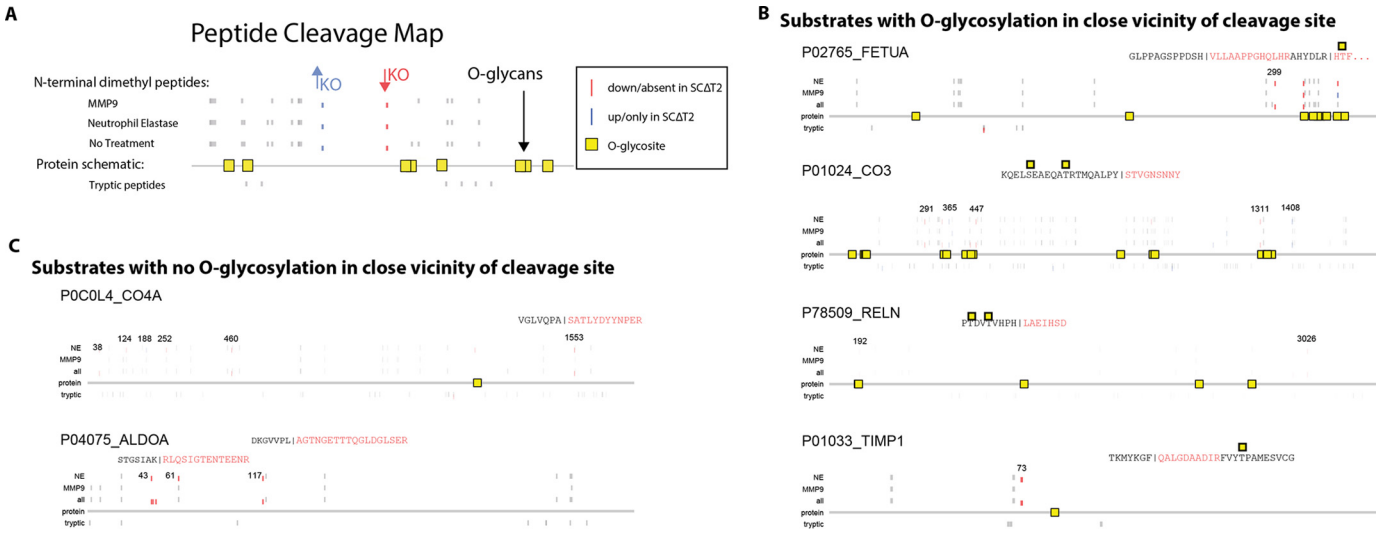


**Figure 4. Cleavage maps and the directional effect of loss of *O*-glycosylation.** *A*, schematic of a peptide map illustrating the location and quantitation of TAILS and pre-TAILS peptides. Peptide maps provide a visual summary of the TAILS data by mapping identified peptides onto schematics of each protein. The *gray bar along* the *x axis* represents the total protein length of the canonical protein isoform, and peptides identified and quantified in TAILS are represented by a *bar* mapped to the P1′ site on the protein (note that peptides are all drawn to the same length). Tryptic peptides or pre-TAILS peptides are shown *below* the *gray bar* and provide a reference point by illustrating any changes in the nonenriched tryptic peptides. *O*-Glycans are indicated by a *yellow square* on the protein schematic. Peptides with *blue shading* have significantly increased cleavage in SCΔT2 cells (category I or II), and peptides with *red shading* are significantly less cleaved in SCΔT2 cells (category IV or V). *Numerical label*, P1′ site coordinate (positioned at the N terminus of the site). For clarity, where multiple overlapping peptides are identified, only the most N-terminal P1′ site is labeled. The *MMP9* and *neutrophil elastase rows* show changes identified within (but not necessarily specific to) these treatment groups. The *all row* illustrates treatment-independent changes, in which data are reanalyzed as an average across all treatment groups. *B*, cleavage maps of four different proteins where cleavage events are detected in close proximity to known *O*-glycosylation sites. The changes in cleavage are potentially due to a direct effect of the nearby *O*-glycan(s). *C*, cleavage maps are shown for two examples, where the observed changes in cleavage are probably due to indirect effects of *O*-glycosylation, as no known *O*-glycosylation sites are located in close proximity to the cleavage site(s).

ment with this finding, pre-TAILS peptides did not show significant changes in total protein (see "tryptic" peptides in cleavage maps in Figs. 4 and 5 and supporting material). Examples of high-confidence targets are shown on peptide maps in Fig. 5*A*. These examples illustrate the occurrence of cleavage events and proximal, distal, and overlapping *O*-glycosites. In

**Table 2**

**High-confidence endogenous cleavage sites significantly altered in SCΔT2 cells**

By substrate winnowing high-confidence targets was identified. 62 cleavage sites in 50 substrates were identified as alternately processed in SCΔT2 cells. For each protein the following information is provided in the respective columns: P1′ position, primary sequence surrounding the cleavage site, median quantification of T2 KO/WT ratio, number of times the peptide was quantified, and known glycosylation status (where Y is yes and N is no). Residues in boldface red type are known *O*-glycosites, and red stretches are reported to be *O*-glycosylated, but without the exact site known. The underlining denotes the sequence of individual peptides identified.

| Entry name | gene | UniProt | P1′ | P1 | | P1′ |
|---|---|---|---|---|---|---|
| A2MG | A2M | P01023 | 587 | NKVDLSFSPSQSLPASHAHLR | \| | VTAAPQSVCALRAVDQSVLLMK |
| AACT | SERPINA3 | P01011 | 25 | MLPLLALGLLAAGFCPAVLCH | \| | PNSPLDEENLTQENQDRGTHVD |
| ALDOA | ALDOA | P04075 | 44 | IVAPGKGILAADESTGSIAKR | \| | LQSIGTENTEENRRFYRQLLLT |
| AMBP | AMBP | P02760 | 186 | FRVVAQGVGIPEDSIFTMADR | \| | GECVPGEQEPEPILIPRVRRAV |
| AMBP | AMBP | P02760 | 335 | LFPYGGCQGNGNKFYSEKECR | \| | EYCGVPGDGDEELLRFSN |
| ANGT | AGT | P01019 | 185 | VPWKDKNCTSRLDAHKVLSAL | \| | QAVQGLLVAQGRADSQAQLLLS |
| APOE | APOE | P02649 | 122 | TPVAEETRARLSKELQAAQAR | \| | LGADMEDVCGRLVQYRGEVQAM |
| APOE | APOE | P02649 | 199 | AVYQAGAREGAERGLSAIRER | \| | LGPLVEQGRVRAATVGSLAGQP |
| B4GA1 | B4GAT1 | O43505 | 65 | QYFEFFPPSPRSVDQVKAQLR | \| | TALASGGVLDASGDYRVYRGLL |
| BGH3 | TGFBI | Q15582 | 534 | NRFSMLVAAIQSAGLTETLNR | \| | EGVYTVFAPTNEAFRALPPRER |
| CALL3 | CALML3 | P27482 | 113 | VFDKDGNGFVSAAELRHVMTR | \| | LGEKLSDEEVDEMIRAADTDGD |
| CBPE | CPE | P16870 | 440 | YKLTASAPGYLAITKKVAVPY | \| | SPAAGVDFELESFSERKEEEKE |
| CD109 | CD109 | Q6YHK3 | 670 | DYIDGVYDNAEYAERFMEENE | \| | GHIVDIHDFSLGSSPHVRKHFP |
| CO3 | C3 | P01024 | 291 | FVIFGIQDGEQRISLPESLKR | \| | IPIEDGSGEVVLSRKVLLDGVQ |
| CO3 | C3 | P01024 | 447 | TKKQELSEAEQATRTMQALPY | \| | STVGNSNNYLHLSVLRTELRPG |
| CO4A | C4A | P0C0L4 | 38 | SLQKPRLLLFSPSVVHLGVPL | \| | SVGVQLQDVPRGQVVKGSVFLR |
| CO4A | C4A | P0C0L4 | 460 | QTISELQLSVSAGSPHPAIAR | \| | LTVAAPPSGGPGFLSIERPDSR… |
| CPVL | CPVL | Q9H3G5 | 434 | KAEKKVWKIFKSDSEVAGYIR | \| | QAGDFHQVIIRGGGHILPYDQP |
| EF1A1 | EEF1A1 | P68104 | 55 | IDKRTIEKFEKEAAEMGKGSF | \| | KYAWVLDKLKAERERGITIDIS |
| EMAL4 | EML4 | Q9HC35 | 143 | PQGQREKKEESHSNDQSPQIR | \| | ASPSPQPSSQPLQIHRQTPESK |
| FAS | FASN | P49327 | 374 | WAPNLHFHSPNPEIPALLDGR | \| | LQVVDQPLPVRGGNVGINSFGF |
| FAS | FASN | P49327 | 1405 | VSLRLVGLKKSFYGSTLFLCR | \| | RPTPQDSPIFLPVDDTSFRWVE |
| FETUA | AHSG | P02765 | 299 | PPSPPLGAPGLPPAGSPPDSH | \| | VLLAAPPGHQLHRAHYDLRHTF |
| FETUA | AHSG | P02765 | 318 | SHVLLAAPPGHQLHRAHYDLR | \| | HTFMGVVSLGSPSGEVSHPRKT |
| FGL1 | FGL1 | Q08830 | 127 | AEFSVYCDMSDGGGWTVIQRR | \| | SDGSENFNRGWKDYENGFGNFV |
| FIBA | FGA | P02671 | 272 | LTDMPQMRMELERPGGNEITR | \| | GGSTSYGTGSETESPRNPSSAG |
| FINC | FN1 | P02751 | 939 | VTDVKVTIMWTPPESAVTGYR | \| | VDVIPVNLPGEHGQRLPISRNT |
| FINC | FN1 | P02751 | 2335 | GAICSCTCFGGQRGWRCDNCR | \| | RPGGEPSPEGTTGQSYNQYSQR |
| FPPS | FDPS | P14324 | 93 | NSDVYAQEKQDFVQHFSQIVR | \| | VLTEDEMGHPEIGDAIARLKEV |
| FUS | FUS | P35637 | 398 | RGGGNGRGGRGRGGPMGRGGY | \| | GGGGSGGGGRGGFPSGGGGGGG… |
| GPC3 | GPC3 | P51654 | 359 | AGKLTTTIGKLCAHSQQRQYR | \| | SAYYPEDLFIDKKVLKVAHVEH |
| GRP78 | HSPA5 | P11021 | 61 | SCVGVFKNGRVEIIANDQGNR | \| | ITPSYVAFTPEGERLIGDAAKN |
| HS90A | HSP90AA1 | P07900 | 52 | SLIINTFYSNKEIFLRELISN | \| | SSDALDKIRYESLTDPSKLDSG |
| IBP1 | IGFBP1 | P08833 | 86 | LPLGAACGVATARCARGLSCR | \| | ALPGEQQPLHALTRGQGACVQE |
| IPSP | SERPINA5 | P05154 | 110 | MQILEGLGLNLQKSSEKELHR | \| | GFQQLLQELNQPRDGFQLSLGN |
| LAMC1 | LAMC1 | P11047 | 590 | QVLSYGQNLSFSFRVDRRDTR | \| | LSAEDLVLEGAGLRVSVPLIAQ |
| LG3BP | LGALS3BP | Q08380 | 443 | SWSARKSQLVYQSRRGPLVKY | \| | SSDYFQAPSDYRYYPYQSFQTP |
| MA1A1 | MAN1A1 | P33908 | 170 | DILLEKKKVAQDQLRDKAPFR | \| | GLPPVDFVPPIGVESREPADAA |
| MATN3 | MATN3 | O15232 | 51 | PVARPGFRRLETRGPGGSPGR | \| | RPSPAAPDGAPASGTSEPGRAR |

**Table 2—*continued***

| Entry name | gene | UniProt | P1′ | P1 | P1′ |
|---|---|---|---|---|---|
| MATN3 | MATN3 | O15232 | 109 | IDSSRSVRPLEFTKVKTFVSR | IIDTLDIGPADTRVAVVNYAST |
| NID1 | NID1 | P14543 | 1057 | FWTDSNLDRIEVAKLDGTQRR | VLFETDLVNPRGIVTDSVRGNL |
| PCSK9 | PCSK9 | Q8NBP7 | 47 | AGARAQEDEDGDYEELVLALR | SEEDGLAEAPEHGTTATFHRCA |
| PDIA5 | PDIA5 | Q14554 | 24 | RAGPAWLLLAIWVVLPSWLSS | AKVSSLIERISDPKDLKKLLRT |
| PLTP | PLTP | P55058 | 18 | MALFGALFLALLAGAHA | EFPGCKIRVTSKALELVKQEGL |
| PROF1 | PFN1 | P07737 | 76 | DRSSFYVNGLTLGGQKCSVIR | DSLLQDGEFSMDLRTKSTGGAP |
| PRSS8 | PRSS8 | Q16651 | 31 | GQLGAVAILLYLGLLRSGTGA | EGAEAPCGVAPQARITGGSSAV |
| RELN | RELN | P78509 | 2936 | CTECGILAEDTALYFGGSTVR | QAVTQDLDLRGAKFLQYWGRIG |
| RELN | RELN | P78509 | 3026 | YILLPEDALTNTTRLRWWQPF | VISNGIVVSGVERAQWALDNIL |
| ROA1 | HNRNPA1 | P09651 | 18 | MSKSESPKEPEQLRKLF | IGGLSFETTDESLRSHFEQWGT |
| ROA2 | HNRNPA2B1 | P22626 | 25 | TLETVPLERKKREKEQFRKLF | IGGLSFETTEESLRNYYEQWGK |
| ROA2 | HNRNPA2B1 | P22626 | 191 | VLQKYHTINGHNAEVRKALSR | QEMQEVQSSRSGRGGNFGFGDS |
| ROA3 | HNRNPA3 | P51991 | 243 | NFMGRGGNFGGGGGNFGRGGN | FGGRGGYGGGGGGSRGSYGGGD |
| TBB6 | TUBB6 | Q9BUF5 | 63 | LQLERINVYYNESSSQKYVPR | AALVDLEPGTMDSVRSGPFGQL |
| THRB | F2 | P00734 | 354 | EADCGLRPLFEKKSLEDKTER | ELLESYIDGRIVEGSDAEIGMS |
| TIMP1 | TIMP1 | P01033 | 73 | VNQTTLYQRYEIKMTKMYKGF | QALGDAADIRFVYTPAMESVCG |
| TRFE | TF | P02787 | 240 | GDVAFVKHSTIFENLANKADR | DQYELLCLDNTRKPVDEYKDCH |
| TRFE | TF | P02787 | 18 | MRLAVGALLVCAVLGLC | LAVPDKTVRWCAVSEHEATKCQ |
| UBA1 | UBA1 | P22314 | 351 | HIGFQALHQFCAQHGRPPRPR | NEEDAAELVALAQAVNARALPA |
| UGDH | UGDH | O60701 | 178 | LSNPEFLAEGTAIKDLKNPDR | VLIGGDETPEGQRAVQALCAVY |
| UGDH | UGDH | O60701 | 374 | ISKYLMDEGAHLHIYDPKVPR | EQIVVDLSHPGVSEDDQVSRLV |
| VILI | VIL1 | P09327 | 506 | PPHLMSIFKGRMVVYQGGTSR | TNNLETGPSTRLFQVQGTGANN |
| XRCC6 | XRCC6 | P12956 | 219 | IFLDLMHLKKPGGFDISLFYR | DIISIAEDEDLRVHFEESSKLE |

contrast to the majority of detected cleavage events, three proteins (*i.e.* serrotransferrin, heat shock protein 90-$\alpha$, and prothrombin) exhibited increased cleavage in SCΔT2 (Fig. 5*B*). The two cleavage sites ($^{192}$**T**VAM$^{196}$**T**PR ↓ $^{199}$SEG ↓ $^{202}$SSV) (where bold highlights possible glycosylation site) identified on prothrombin were found to be closely flanked by *O*-glycosylation sites at Thr$^{192}$ and Thr$^{196}$. Cleavage at **T**PR ↓ $^{199}$SEG is performed by thrombin and is associated with altered prothrombin activation (51, 52), and we have previously demonstrated that *O*-glycosylation in this region alters the sensitivity to thrombin in peptidic assays (38). Interestingly, GalNAc-T2–specific glycosites in thrombin have previously been detected from both human and rodent samples (25), further suggesting a role for GalNAc-T2 in thrombin function. A similar protection from cleavage was also evident around signal peptide sites, with a surprising number of changes detected in this region (Fig. 5*C*). Importantly, these changes are not due to loss of a protective effect at the signal peptide processing site, as there is no evidence of *O*-glycosylation occurring in this region. Cleavage maps summarizing GalNAc-T2– dependent quantitative changes per protein are available in Fig. S1.

### *GalNAc-T2–induced changes in regulators of lipid homeostasis*

Loss of GalNAc-T2 has recently been associated with decreased levels of high-density lipoprotein cholesterol (25). It has been shown that GalNAc-T2 directly glycosylates Apolipoprotein C-III (apoCIII), angiopoietin-related protein 3 (ANGPTL3), phospholipid transfer protein, and others. This glycosylation is proposed to affect the function of these proteins and thereby regulate HDL levels. For ANGPTL3, it was demonstrated that glycosylation inhibited proteolytic cleavage (29, 43), and we therefore hypothesized that we could use TAILS to detect ANGPTL3 and to find evidence for GalNAc-T2 glycosylation of additional regulators of lipid homeostasis. Whereas neither apoC-III nor ANGPTL3 peptides were identified in our data set, signal peptide cleavage of phospholipid transfer protein was increased in SCΔT2 cells. Using TAILS, we were also able to monitor the processing of several major apolipoproteins. No substantial changes were detected in apoA1. However, we found processing of the central repeat regions of apoE and apoH to be altered by loss of GalNAc-T2 at low frequency (~1% of total cleavage events). In agreement, processing did not differ between SC and SCΔT2 for either apoA1 or apoE

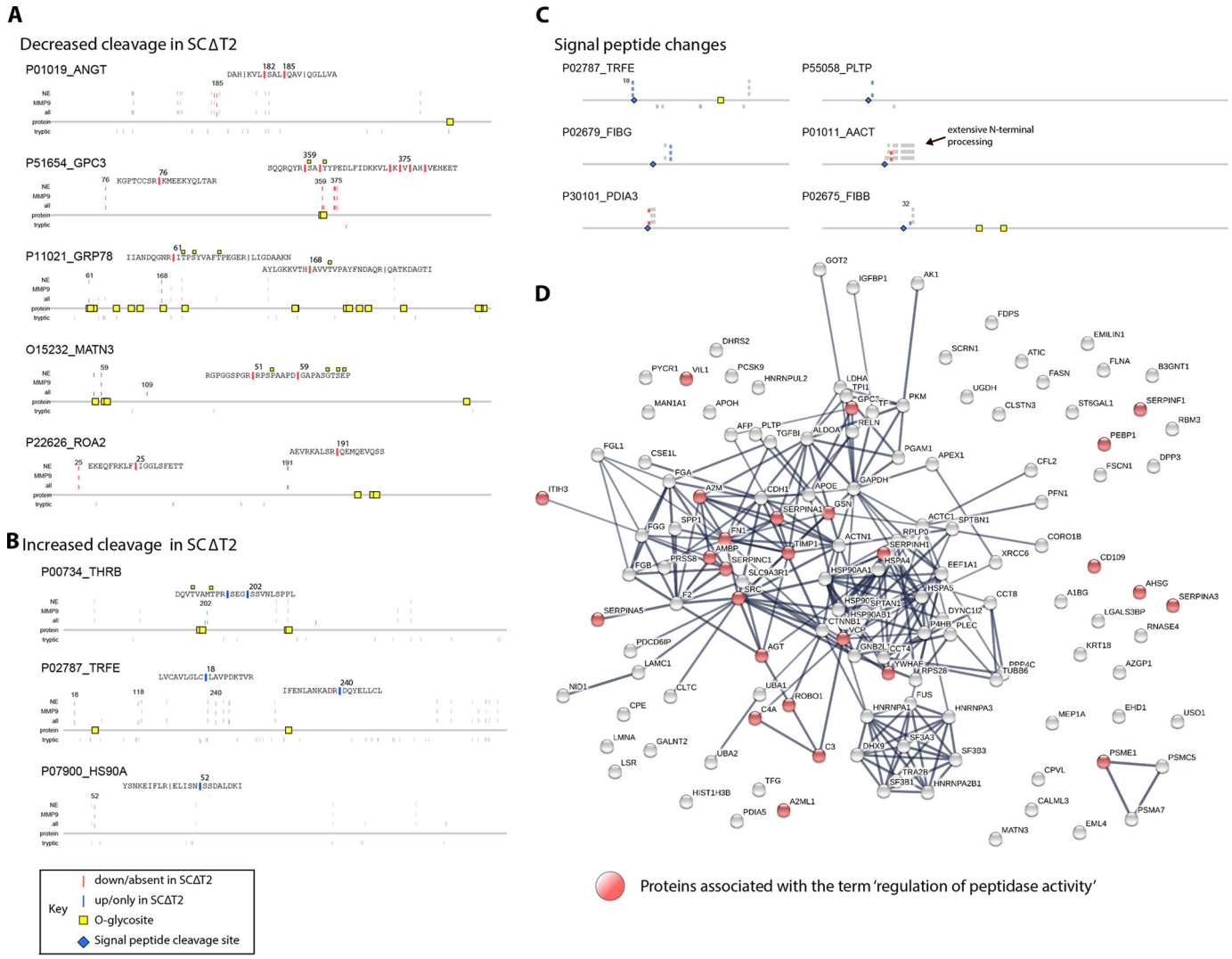


**Figure 5. Cellular protein processing is altered with loss of GALNT2.** *A*, examples of proteins in which deletion of GALNT2 results in decreased cleavage independent of treatment with exogenous proteases. *B*, examples illustrating some of the few proteins for which loss of GALNT2 results in increased cleavage. *C*, modified peptide maps of the N-terminal 70 amino acids of proteins in which changes in signal peptide processing were observed. In contrast to the remainder of the cleavage events, loss of GALNT2 results in increased signal peptide cleavage for most of these proteins. *D*, relaxed *K*-means clustered network of proteins displaying altered proteolytic processing with loss of GALNT2. The network was generated using StringDB. Proteins associated with the term "regulation of peptidase activity" are *highlighted* in *red*.

when assayed by Western blotting (data not shown). This suggests that GalNAc-T2 is a minor regulator of proteolytic processing of these proteins, potentially due to the distance of the modification from the cleavage site. Conversely, substantial changes were detected in the C-terminal region of apoB, for which there was a marked increase in tryptic peptides from SCΔT2 supernatant from $^{2339}$AKV. This indicates that loss of GalNAc-T2 alters the stability of the apoB C terminus. The apoB C terminus is required for binding to the low-density lipoprotein receptor and is therefore a major mediator of HDL uptake (53). Furthermore, decreased cleavage of lipolysis-stimulated lipoprotein receptor, fatty acid synthase, and zinc-α-2-glycoprotein (ZA2G) was found in in SCΔT2 cells. Such altered processing of lipid-regulating proteins may contribute to the decrease in high-density lipoprotein cholesterol in a number of mammalian species carrying mutations in *GALNT2* (25).

### *GALNT2 deletion perturbs the protease web*

To decipher why we observed a significant number of cleaved endogenous N termini peptides that were altered in SCΔT2, we performed gene ontology (GO) term analysis on proteins containing significantly changed N termini. Strikingly, the five most significant GO terms were associated with the regulation of peptidase activity. Of the 26 proteins designated as regulators of proteolysis, six were serpins. Indeed, serpin domains were also found to be significantly enriched in a separate analysis of protein domains (FDR = $1.46 \times 10^{-6}$). We therefore hypothesized that dysregulation of serpins, some of which are key connectors of nodes within the protease web (54), may explain many of the changes identified in the endogenous N-terminome. To determine whether this was the case, we used the STRINGdb with K-means clustering to identify whether there were potential interactions between alternately cleaved proteins (Fig. 5*E*). The network was composed of 132 nodes with

243 edges, representing a significant enrichment of interactions (clustering coefficient 0.701, $p \ll 0.001$). Cluster analysis identified four potential functional groups formed by proteins involved in the unfolded protein response, glycolysis, and extracellular matrix organization. Interestingly, all clusters contained regulators of proteolysis, and the greatest number of connections were found for the three serpins antithrombin III, serpin H1, and $\alpha$-1-antitrypsin, as well as $\alpha_2$-macroglobulin, and the metalloproteinase inhibitor TIMP1, suggesting that these proteins are key regulators of the GalNAc-T2– dependent response.

## Discussion

*O*-Glycosylation has previously been hypothesized to generally co-regulate proteolytic processing by blocking cleavage (28, 29, 55). This is based on a number of studies with selected examples of site-specific *O*-glycosylation blocking specific cleavage sites (20–23, 25, 27, 43, 55, 56). Furthermore, classical studies have established that dense *O*-glycosylation provides protection and can confer stability to cell-surface proteins (19, 57). In the current study, we have taken an unbiased and system-wide approach to test the interplay between *O*-glycosylation and proteolytic cleavage. Surprisingly, we find that loss of GalNAc-T2 activity in a SimpleCell predominantly leads to decreased cleavage of proteins (Fig. 3*D*), whereas for a small number of substrates, the cleavage was increased. Furthermore, we observe changes in several central players of the protease web, including proteases and protease inhibitors, suggesting that loss of *O*-glycosylation leads to general changes in the protease network and hence changes in the proteolytic potential of the system.

In our study, we applied two orthogonal methods: an *in silico* screen and a glyco-degradomics strategy. Both approaches showed evidence for cross-talk and indicated that the effect of glycosylation was specific to the protease under investigation. From the *in silico* screen, we found that up to 35% of protease cleavage sites from the MMP protease family occurred in *O*-glycosylated regions. Furthermore, by TAILS, we found that loss of GalNAc-T2 impacted protein cleavage. However, surprisingly, loss of GalNAc-T2 also resulted in decreased cleavage with a ~13% decrease of endogenous N termini and a 32% decrease of the neutrophil elastase N-terminome without changing overall protein stability. Together, these data demonstrate co-localization of *O*-glycosylation and protease cleavage and show that changes in *O*-glycosylation markedly impact the N-terminome. The *in silico* model and experimental data are in agreement with prior *in vitro* studies, in which the closer the *O*-glycan resides to the scissile bond, the greater the impact on cleavage (26). Furthermore, our findings demonstrate that *O*-glycans are systematically enriched around verified MMP and certain serine protease cleavage sites (Fig. 1). Based on earlier studies demonstrating that *O*-glycans mostly inhibit cleavage, we expected increased proteolytic cleavage with loss of site-specific *O*-glycosylation due to knockout of GALNT2, yet we unexpectedly also observed a decrease in cleavage. However, increased processing of glycosylated substrates is not without precedent, as demonstrated for ADAM proteases using *in vitro* peptide assays (28, 58, 59). Thus, it is possible that *loss* of *O*-glycosylation close to cleavage sites in some cases *decreases* cleavage and that this is a more common mechanism than previously understood.

The driving influence of protease exosites, substrate-binding sites outside the active site and often on ancillary exosite domains (60), markedly improves the $k_{cat}/K_m$ for many substrates. With our examples of enhanced cleavage in the vicinity of glycosites, proteases may bind sugar groups as a new class of glyco-exosite that can be regulated.

Another possibility is that the observed decrease in cleavage despite loss of GalNAc-T2 could be driven by perturbations in the protease network, caused by loss of *O*-glycosylation on proteins that are not substrates but regulators of proteolytic activity. The GO term analysis further favors this explanation, identifying peptidase activity as the most important term for the changed N termini.

Several technical limitations need to be taken into consideration. First, it should be noted that *O*-glycopeptide signals are often lost during MS analysis due to their low abundance and relatively poor ionization efficiencies (61). The simple presence of a glycan will therefore alter the KO/WT ratio in cases where an *O*-glycan resides with high occupancy (>60%) on the identified neo-N-terminal peptide (*i.e.* on the prime side of the scissile bond). These cases must be carefully validated to establish that change in KO/WT is in fact due to change in proteolytic cleavage. In our TAILS experiments, these cases occur on ~16% of the significantly altered peptides; thus, it is unlikely that this accounts for any major part of the observed changes. Second, to reduce the complexity of MS analysis, we employed a SimpleCell approach. This ensures that all *O*-glycans are truncated, generating a more homogeneous *O*-glycoproteome. Therefore, we cannot at this stage exclude the possibility that glycan elongation will have additional or even opposite effects on the proteolytic cleavage events observed. Using TAILS to monitor changes in the N-terminome allows precision monitoring of cleavage events across a protein. The advantage of such a technique is that the combined effects of the protease web can be assessed in the cellular context. However, this also brings challenges in deconvoluting the underlying mechanisms. In particular, it can be difficult to distinguishing between changes due to loss of close-by glycosylation *versus* more indirect effects, such as perturbations in the protease network. Furthermore, our understanding of the *O*-glycoproteome is still hampered by lack of information of occupancy regarding individual *O*-glycosites; therefore, it is possible that we are overestimating the occupancy of a number of *O*-glycosites. Consequently, our *in silico* analysis, demonstrating significant overlap between protease cleavage sites and detected glycosylation, represents a possible rather than an established co-regulation between *O*-glycosylation sites and cleavage sites. In this context, it is important to note that GalNAc-Ts controlling *O*-glycosylation sites are differentially expressed in a temporal and spatial manner, and the occupancy of individual sites is therefore predicted to vary, not least for sites with potential regulatory functions (33). Validation of individual targets is therefore important but may not be trivial. Although low-occupancy glycosylation sites can be of significant biological importance (62), they can be difficult to detect due to masking by the nonglyco-

sylated pool of that protein and should therefore be addressed in relevant biological settings. Notwithstanding these caveats and technical limitations, our study uncovered a number of compelling findings and demonstrates that TAILS can screen for interactions between these PTMs and identification of novel GalNAc-T2 targets, and the data presented here will serve as a valuable entry point for follow-up studies of individual targets.

Importantly, TAILS aided in the prioritization of relevant GalNAc-T2 targets affected by processing, which has previously been difficult to achieve. Functionally relevant *O*-glycosites are difficult to identify due to the lack of information regarding stoichiometry of glycan occupancy. TAILS partly circumvents this issue, as it provides a quantitative readout for individual cleavage events. Indeed, using TAILS, we identified a number of processing events in regulators of lipid metabolism affected by GalNAc-T2 glycosylation. One identified target of TAILS was reelin. Reelin is a secreted glycoprotein that binds VLDLR and the APOER2 receptors with high affinity; it is required for correct neuronal migration and is implicated in the regulation of hemostasis and atherosclerosis (63–65). Reelin was previously identified as a GalNAc-T2 target in human plasma (25), and we found that cleavage of reelin at the N-terminal $^{184}$***T***DV$^{187}$***T***VHPH ↓ $^{192}$LAE and C-terminal QPF ↓ $^{3026}$VIS sites was decreased in SCΔT2 cells. Cleavage at these sites resulted in loss of the N-terminal reeler domain and the conserved C-terminal domains from the central protein, which are required for downstream signaling in tissues (66, 67). It is therefore of interest to determine whether the function of reelin is in fact co-regulated by GalNAc-T2 *in vivo*.

In conclusion, our data demonstrate the occurrence of a complex interplay between *O*-glycosylation and proteolytic cleavage, including cleavage events probably affected by nearby glycosylation as well as broad changes in the protease web. It is possible that *O*-glycans and proteases interact on multiple levels, as suggested by the extensive changes in the protease network of GalNAc-T2 KO SC cells. Further elucidation of the underlying mechanisms will require focused studies on target proteins and structural characterization of the impact of *O*-glycosylation on protease-substrate recognition and binding. Furthermore, much remains to be understood about the potential impact of this interaction. Whereas we detect differences in the response to neutrophil elastase, studies on additional proteases are required to fully understand the relationship between *O*-glycosylation, substrate specificity, and catalytic mechanism. Improved methods for validating candidate targets identified in TAILS screens are also required, as biochemical assays using recombinant proteins are not necessarily representative of protein processing in the cellular context, where surface charge and cofactors often play important roles, and design of the requisite assays is not feasible for a large number of candidate substrates. In summary, we show that utilization of *in silico* tools and TAILS can aid in the identification of cleavage events regulated by site-specific *O*-glycosylation. Our findings highlight that we are still far from truly understanding the co-regulatory role of *O*-glycosylation in proteolytic processing events, yet these results demonstrate that global approaches provide important information about the complexity of these systems and that such studies are needed to fully appreciate the interplay between these two PTMs. In addition, we have found an unexpected increase in some cleavage events associated with nearby glycans, which indicates complex binding interactions between substrate and protease to favor cleavage.

## Experimental procedures

### *In silico screen*

All analysis was performed using R. GalNAc-type *O*-glycosylation and protease cleavage site data were obtained from the glycodomain (68) and TopFIND (69) databases, respectively. Analysis was only performed on the canonical protein isoform as defined in the UniProtKB database. Cleavage sites for which the effector protease has not been described (those labeled as "nterm" in the TopFIND database) were discarded from the data set. Only proteases with well-defined substrate degradomes (>20 proteins, >40 sites) were included for further analysis. Disordered regions were defined by extracting disordered loops/coil predictions from the DisEMBL prediction server (70). All analysis was restricted to human and murine species; no appreciable difference was noted when analysis was performed on a single species compared with when site data from both species were analyzed simultaneously.

### *Consensus sequence analysis*

To assess whether *O*-glycosites were under- or overenriched, amino acid sequences were obtained for each protein in the filtered TopFIND data set, and known *O*-glycosylated Ser or Thr amino acids were recoded as *X*, thereby treating these modified residues as independent amino acids. For generation of the reference data set, 10 psuedo-cleavage sites were randomly sampled from each protein listed in the filtered TopFIND data set. The sequence window ± 7 amino acids from P1′ was identified using the canonical sequence from the UniProtKB database, and cleavage sites <21 amino acids from the N or C terminus of proteins were removed. As both *O*-glycans and proteolytic cleavage sites tend to residue in disordered regions, analysis was performed both with and without disorder correction. For disorder correction, only reference and experimentally derived cleavage sites residing within disordered loops/coils were considered. IceLogo plots and heat maps were generated using IceLogo version 2.1 (71), with a $p$ value threshold of 0.001. Collagen and elastin substrates were removed from the MMP analysis, as the majority of these data originate from a series of focused studies, and the homogeneity of these target sequences resulted in the presence of confounding and highly dominant repetitive glycine residues. Due to the highly similar consensus sequences within protease subfamilies, proteases were analyzed as groups. Small differences in protease specificity were noted between highly homologous proteases; however, dominant trends were stable irrespective of the analysis approach.

### *Glycoengineered cell models*

The hepatocellular carcinoma cell line HepG2 was used for all experiments. Cells were glycoengineered by targeted gene editing using zinc-finger nucleases as described previously (24).

### TAILS

Cells were cultured to 80% confluence in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Cells were washed extensively with PBS and grown for a subsequent 24 h in phenol red–free and serum-free medium. Conditioned medium was harvested, and protease inhibitor was immediately added (1 mM phenylmethylsulfonyl fluoride). The medium was cleared of cell debris at 2,200 × *g* for 5 min, followed by vacuum filtration through a 0.45-μm membrane, and proteins were concentrated using Amicon Ultra-15 centrifugal filters (3-kDa cutoff; Millipore). Cell lysates were harvested into 0.5 mM EDTA/PBS with 1 mM phenylmethylsulfonyl fluoride and pelleted at 500 × *g* for 15 min. Cell pellets were resuspended in 100 mM HEPES, pH 7.0, with protease inhibitor and lysed using a probe sonicator. Lysates were subsequently cleared at 4,200 × *g* for 30 min, followed by filtration through a 0.45-μm syringe filter. Both concentrated medium and cell lysates were subsequently dialyzed five times through Amicon Ultra-15 centrifugal filters (3-kDa cutoff; Millipore) using 100 mM HEPES, pH 7.0, to remove free amino acids (as a source of primary amines). Protein concentration was then measured by a bicinchoninic acid assay (Pierce), and samples were adjusted to 1 or 2 mg/ml. All procedures were performed at 4 °C, and resulting samples were stored at −80 °C until further use. For protease cleavage, 1 mg of total protein was treated with 4-aminophenylmercuric acetate–activated MMP9 (1:100 w/w) or neutrophil elastase (1:500 w/w), in HEPES buffer, pH 7.0, supplemented with 150 mM NaCl and 10 mM $CaCl_2$ for 16 h at 37 °C.

### N-terminal peptide enrichment

Protease-treated samples were immediately prepared for N-terminal enrichment as follows. First, samples were denatured in 4 M guanidine hydrochloride at pH 8.0, reduced with 5 mM DTT for 1 h at 65 °C, and subsequently alkylated with 15 mM iodoacetamide for 30 min at room temperature in the dark. Samples were then adjusted to pH 6.5 for optimal dimethylation. Primary amine groups were labeled by the addition of 40 mM $^{12}CH_2$-formaldehyde (light, +28 Da) (Sigma) or $^{13}CD_2$-formaldehyde (heavy, +34 Da) and 20 mM $NaBH_3CN$. Reactions were incubated overnight at 37 °C and then quenched with 100 mM ammonium bicarbonate (pH 7.0) for 1 h at 37 °C. Excess cyanoborohydride was then decomposed by acidification below pH 2.5 using TFA and incubation for 1 h at 37 °C. Samples were then combined and precipitated by chloroform-methanol precipitation and reconstituted into 50 mM HEPES, pH 8. Proteins were then prepared for shotgun analysis by overnight digestion with sequencing grade trypsin (Roche Applied Science; 1:50 (w/w)) at 37 °C. Complete digestion was confirmed by 4–12% SDS-PAGE and Coomassie staining, and a pre-TAILS aliquot was saved for subsequent LC-MS analysis of total protein. Finally, TAILS enrichment was performed by negative selection using a hyperbranched polyglycerol polyaldehyde-derivatized polymer (HPG-ALDII, from Flintbox, track code 08-038) to covalently bind free primary amines generated during tryptic digest. The samples were first adjusted to pH 6–7, 20 mM $NaBH_3CN$ was added, and then samples were incubated overnight with a 5-fold excess of HPG-ALDII polymer to peptide (w/w) to ensure complete coupling of all trypsin-generated, internal peptides. Unbound peptides were then separated from the high-molecular weight polymer by filtration through a 3-kDa centrifugal filter, and when concentrated to ~100 μl, the polymer was further washed using 0.2 ml of 100 mM ammonium bicarbonate, the flow-through fractions were combined, and the polymer with coupled tryptic peptides was discarded. The eluted free N termini were desalted and concentrated on C-18 StageTips for MS analysis.

### In-line liquid chromatography and mass spectrometry analysis

EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via nanoSpray Flex ion source to an Orbitrap Fusion MS or LTQ-Orbitrap Velos Pro spectrometer (Thermo Scientific) was used for analysis as described previously (72, 73). Briefly, the nLC was operated in a single analytical column set up using PicoFrit Emitters (New Objectives, 75-μm inner diameter) packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, 1.9-μm particle size, 19–21-cm column length). Each sample was injected onto the column and eluted in a gradient from 2 to 20% B in 95 min, from 20% to 80% B in 10 min, and 80% B for 15 min at 200 nl/min (solvent A, 100% $H_2O$; solvent B, 100% acetonitrile; both containing 0.1% (v/v) formic acid). A precursor MS1 scan ($m/z$ 350–1,700) of intact peptides was acquired in the Orbitrap at the nominal resolution setting of 30,000 for Velos Pro and 120,000 for Fusion, followed by Orbitrap HCD-MS2 of the five most abundant multiply charged precursors in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 was used for triggering data-dependent fragmentation events. Supplemental activation of the charge-reduced species was used in the electron-transfer dissociation analysis to improve fragmentation. A 1-min dynamic exclusion window was used to prevent repeated analysis of the same components.

### Data analysis

Raw data were analyzed using Thermo Scientific Proteome Discoverer version 2.1 software and searched against the UniProt KB/SwissProt–reviewed database downloaded on July 8, 2010, containing 20,212 entries. An additional fasta file containing contaminants obtained from a common repository at the Max Planck Institute was included in the search. HCD data were searched using the SEQUEST HT node in PD 2.1. In all cases, the precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 50 millimass units. Carbamidomethylation (+57,021 Da) of cysteine and dimethyl labeling (light, +28,031 Da; heavy, +34,069 Da) of lysine were set as static modifications. Dimethylation of the protein N termini, acetylation of protein N termini (+42,011 Da), formation of pyroglutamate on peptide N termini (+17,027 Da), and methionine oxidation (+15,996 Da) were considered dynamic modifications. The search was conducted using semispecific trypsin cleavage with up to two missed cleavages. All spectra were searched against nondecoy and decoy databases to allow calculation of the false discovery rate. Afterward, Percolator was used to calculate the local false discovery rate, and a threshold FDR of 1% was applied at the peptide level. Data were subsequently analyzed with R. The technical variance was measured by comparison of two identical nontreated samples (Fig. 3). The samples were differentially labeled and compared using TAILS; in the ideal scenario, quantitation of the two fractions should result in a heavy/light ratio of 1.0 for all quantified N-terminal peptides. A deviation from this ratio represents the technical error associated with TAILS enrich-

ment and quantitation by LC-MS/MS. From this analysis, 98% of peptides were quantified with a <2-fold change (H/L ratio range $\log_2$ −0.73 to 0.21), indicating good technical replication. Subsequently, changes in stability and/or secretion of individual proteins were assayed by comparing the pre-TAILS samples consisting of nonenriched tryptic peptide from the different cell sources. The peptide dimethyl ratios of the internal tryptic peptides were combined by sample type and used to define significant outlier cut-offs by box-and-whisker plot analysis using the BoxPlotR tool (74) as described previously (75) (coefficient = 1). Final $\log_2$ upper and lower threshold values were 1.52 and −1.69, respectively. These results are in agreement with previously reported lower limits of detection for dimethyl quantitation and TAILS experiments (11, 76).

### *Peptide maps*

Data for proteins containing quantified N termini were subsequently mapped onto protein schematics using R to produce a peptide map summarizing total quantitation from all experiments. These barcodes provide a visual method for assessing the changes detected in each treatment group. Two separate analyses were performed to obtain these data; as MMP9– and neutrophil elastase–specific effects were only a small portion of the total data set, the analysis was repeated by combining all experimental groups to generate treatment-independent quantification. These data are represented by the "shotgun" and "all" rows in the peptide map.

### *Substrate winnowing*

Correctly mapped peptides designated as outliers were subjected to a substrate-winnowing process to separately identify protease-dependent and -independent targets. Only peptides identified in at least two spectra were considered, and peptides identified with variable N termini forms (*e.g.* a peptide identified with both an N-terminal dimethyl and N-terminal pyroglutamate label in the same sample) were excluded due to the uncertainty in quantitation of these species. In cases where multiple spectra support the quantification of a single species of peptide, quantifications were merged by summing the total quantification areas (as determined by Proteome Discoverer) for each channel, and the ratio of the two was used as the quantification ratio (Fig. S2). Protease-independent targets were defined as outlier peptides identified across multiple treatment groups, whereas protease-dependent targets were defined as high-confidence targets uniquely identified in that treatment group. Peptides quantified only in one channel were only accepted for further consideration if they were identified by at least two high-quality spectra. Such peptides were deemed singlets and assigned a ratio of 1:50 based on the channel in which they were identified. Of the remaining peptides, ~50% of treatment-independent and 65% of treatment-dependent (MMP9 or neutrophil elastase) substrates were known *O*-glycoproteins. Substrates passing these quality checks are listed in the candidate peptide tables (Tables 1 and 2). GO term analysis of protease substrates was performed using the DAVID (78), Gorilla (79), and REVIGO (80) tools. Network effects were assessed with STRING (81). All analysis was performed using unchanged proteins identified in HepG2 cells during TAILS as the reference data set.

*Author contributions*—S. L. K., C. K. G., U. E., K. T. S., C. M. O., and H. H. W. conceptualization; S. L. K., C. K. G., U. E., and H. J. J. formal analysis; S. L. K., C. K. G., U. E., A. D. H., and S. Y. V. investigation; S. L. K., C. K. G., U. E., H. J. J., S. Y. V., K. T. S., C. M. O., and H. H. W. methodology; S. L. K., C. K. G., and H. H. W. writing-original draft; S. L. K., C. K. G., U. E., H. J. J., A. D. H., K. T. S., C. M. O., and H. H. W. writing-review and editing; K. T. S., C. M. O., and H. H. W. supervision; C. M. O. and H. H. W. resources.

## References


1. Eckhard, U., Marino, G., Butler, G. S., and Overall, C. M. (2016) Positional proteomics in the era of the human proteome project on the doorstep of precision medicine. *Biochimie* **122,** 110–118 CrossRef Medline
2. Rawlings, N. D., Barrett, A. J., and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **44,** D343–D350 CrossRef Medline
3. Deu, E., Verdoes, M., and Bogyo, M. (2012) New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat. Struct. Mol. Biol.* **19,** 9–16 CrossRef Medline
4. Dufour, A., and Overall, C. M. (2013) Missing the target: matrix metalloproteinase antitargets in inflammation and cancer. *Trends Pharmacol. Sci.* **34,** 233–242 CrossRef Medline
5. Overall, C. M., and Blobel, C. P. (2007) In search of partners: linking extracellular proteases to substrates. *Nat. Rev. Mol. Cell Biol.* **8,** 245–257 CrossRef Medline
6. Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* **5,** 785–799 CrossRef Medline
7. Davie, E. W., and Ratnoff, O. D. (1964) Waterfall sequence for intrinsic blood clotting. *Science* **145,** 1310–1312 CrossRef Medline
8. Weber, S., and Saftig, P. (2012) Ectodomain shedding and ADAMs in development. *Development* **139,** 3693–3709 CrossRef Medline
9. Wen, W., Moses, M. A., Wiederschain, D., Arbiser, J. L., and Folkman, J. (1999) The generation of endostatin is mediated by elastase. *Cancer Res.* **59,** 6052–6056 Medline
10. Bellac, C. L., Dufour, A., Krisinger, M. J., Loonchanta, A., Starr, A. E., Auf dem Keller, U., Lange, P. F., Goebeler, V., Kappelhoff, R., Butler, G. S., Burtnick, L. D., Conway, E. M., Roberts, C. R., and Overall, C. M. (2014) Macrophage matrix metalloproteinase-12 dampens inflammation and neutrophil influx in arthritis. *Cell Rep.* **9,** 618–632 CrossRef Medline
11. auf dem Keller, U., Prudova, A., Eckhard, U., Fingleton, B., and Overall, C. M. (2013) Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation. *Sci. Signal.* **6,** rs2 Medline
12. Overall, C. M., McQuibban, G. A., and Clark-Lewis, I. (2002) Discovery of chemokine substrates for matrix metalloproteinases by exosite scanning: a new tool for degradomics. *Biol. Chem.* **383,** 1059–1066 Medline
13. Wigle, T. J., Provencher, L. M., Norris, J. L., Jin, J., Brown, P. J., Frye, S. V., and Janzen, W. P. (2010) Accessing protein methyltransferase and demethylase enzymology using microfluidic capillary electrophoresis. *Chem. Biol.* **17,** 695–704 CrossRef Medline
14. Turowec, J. P., Zukowski, S. A., Knight, J. D., Smalley, D. M., Graves, L. M., Johnson, G. L., Li, S. S., Lajoie, G. A., and Litchfield, D. W. (2014) An unbiased proteomic screen reveals caspase cleavage is positively and negatively regulated by substrate phosphorylation. *Mol. Cell Proteomics* **13,** 1184–1197 CrossRef Medline
15. Dix, M. M., Simon, G. M., Wang, C., Okerberg, E., Patricelli, M. P., and Cravatt, B. F. (2012) Functional interplay between caspase cleavage and phosphorylation sculpts the apoptotic proteome. *Cell* **150,** 426–440 CrossRef Medline
16. Russell, D., Oldham, N. J., and Davis, B. G. (2009) Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydr. Res.* **344,** 1508–1514 CrossRef Medline

17. Meldal, M., and Bock, K. (1994) A general approach to the synthesis of O- and *N*-linked glycopeptides. *Glycoconj. J.* **11,** 59–63 CrossRef Medline
18. Ihara, S., Miyoshi, E., Nakahara, S., Sakiyama, H., Ihara, H., Akinaga, A., Honke, K., Dickson, R. B., Lin, C. Y., and Taniguchi, N. (2004) Addition of β1–6 GlcNAc branching to the oligosaccharide attached to Asn 772 in the serine protease domain of matriptase plays a pivotal role in its stability and resistance against trypsin. *Glycobiology* **14,** 139–146 Medline
19. Kozarsky, K., Kingsley, D., and Krieger, M. (1988) Use of a mutant cell line to study the kinetics and function of *O*-linked glycosylation of low density lipoprotein receptors. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4335–4339 CrossRef Medline
20. Mains, R. E., and Eipper, B. A. (1981) Differences in the post-translational processing of β-endorphin in rat anterior and intermediate pituitary. *J. Biol. Chem.* **256,** 5683–5688 Medline
21. Herbert, E., Budarf, M., Phillips, M., Rosa, P., Policastro, P., Oates, E., Roberts, J. L., Seidah, N. G., and Chrétien, M. (1980) Presence of a pre-sequence (signal sequence) in the common precursor to ACTH and endorphin and the role of glycosylation in processing of the precursor and secretion of ACTH and endorphin. *Ann. N.Y. Acad. Sci.* **343,** 79–93 CrossRef Medline
22. Birch, N. P., Estivariz, F. E., Bennett, H. P., and Loh, Y. P. (1991) Differential glycosylation of N-POMC1–77 regulates the production of γ 3-MSH by purified pro-opiomelanocortin converting enzyme: a possible mechanism for tissue-specific processing. *FEBS Lett.* **290,** 191–194 CrossRef Medline
23. Kato, K., Jeanneau, C., Tarp, M. A., Benet-Pagès, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., and Clausen, H. (2006) Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis: secretion of fibroblast growth factor 23 requires *O*-glycosylation. *J. Biol. Chem.* **281,** 18370–18377 CrossRef Medline
24. Schjoldager, K. T., Vakhrushev, S. Y., Kong, Y., Steentoft, C., Nudelman, A. S., Pedersen, N. B., Wandall, H. H., Mandel, U., Bennett, E. P., Levery, S. B., and Clausen, H. (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 9893–9898 CrossRef Medline
25. Khetarpal, S. A., Schjoldager, K. T., Christoffersen, C., Raghavan, A., Edmondson, A. C., Reutter, H. M., Ahmed, B., Ouazzani, R., Peloso, G. M., Vitali, C., Zhao, W., Somasundara, A. V., Millar, J. S., Park, Y., Fernando, G., *et al.* (2016) Loss of function of GALNT2 lowers high-density lipoproteins in humans, nonhuman primates, and rodents. *Cell Metab.* **24,** 234–245 CrossRef Medline
26. Schjoldager, K. T., Vester-Christensen, M. B., Goth, C. K., Petersen, T. N., Brunak, S., Bennett, E. P., Levery, S. B., and Clausen, H. (2011) A systematic study of site-specific GalNAc-type *O*-glycosylation modulating proprotein convertase processing. *J. Biol. Chem.* **286,** 40122–40132 CrossRef Medline
27. Goth, C. K., Tuhkanen, H. E., Khan, H., Lackman, J. J., Wang, S., Narimatsu, Y., Hansen, L. H., Overall, C. M., Clausen, H., Schjoldager, K. T., and Petäjä-Repo, U. E. (2017) Site-specific *O*-glycosylation by polypeptide *N*-acetylgalactosaminyltransferase 2 (GalNAc-transferase T2) co-regulates β1-adrenergic receptor N-terminal cleavage. *J. Biol. Chem.* **292,** 4714–4726 CrossRef Medline
28. Goth, C. K., Halim, A., Khetarpal, S. A., Rader, D. J., Clausen, H., and Schjoldager, K. T. (2015) A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific *O*-glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 14623–14628 CrossRef Medline
29. Schjoldager, K. T., and Clausen, H. (2012) Site-specific protein *O*-glycosylation modulates proprotein processing: deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochim. Biophys. Acta* **1820,** 2079–2094 CrossRef Medline
30. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., *et al.* (2013) Precision mapping of the human *O*-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32,** 1478–1488 CrossRef Medline
31. Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-*N*-acetyl-α-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.* **272,** 23503–23514 CrossRef Medline
32. Kong, Y., Joshi, H. J., Schjoldager, K. T., Madsen, T. D., Gerken, T. A., Vester-Christensen, M. B., Wandall, H. H., Bennett, E. P., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2015) Probing polypeptide GalNAc-transferase isoform substrate specificities by *in vitro* analysis. *Glycobiology* **25,** 55–65 CrossRef Medline
33. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type *O*-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22,** 736–756 CrossRef Medline
34. Tian, E., and Ten Hagen, K. G. (2007) O-linked glycan expression during *Drosophila* development. *Glycobiology* **17,** 820–827 CrossRef Medline
35. Ten Hagen, K. G., Tran, D. T., Gerken, T. A., Stein, D. S., and Zhang, Z. (2003) Functional characterization and expression analysis of members of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family from *Drosophila* melanogaster. *J. Biol. Chem.* **278,** 35039–35048 CrossRef Medline
36. Tabak, L. A. (2010) The role of mucin-type *O*-glycans in eukaryotic development. *Semin. Cell Dev. Biol.* **21,** 616–621 CrossRef Medline
37. Luo, S. Z., Mo, X., Afshar-Kharghan, V., Srinivasan, S., López, J. A., and Li, R. (2007) Glycoprotein Ibα forms disulfide bonds with 2 glycoprotein Ibβ subunits in the resting platelet. *Blood* **109,** 603–609 CrossRef Medline
38. King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., and Wandall, H. H. (2017) Characterizing the *O*-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Adv.* **1,** 429–442 CrossRef Medline
39. Gould, W. R., Silveira, J. R., and Tracy, P. B. (2004) Unique *in vivo* modifications of coagulation factor V produce a physically and functionally distinct platelet-derived cofactor: characterization of purified platelet-derived factor V/Va. *J. Biol. Chem.* **279,** 2383–2393 CrossRef Medline
40. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J. N., and Overall, C. M. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* **6,** 1578–1611 CrossRef Medline
41. Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28,** 281–288 CrossRef Medline
42. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., and Clausen, H. (2015) Deconstruction of *O*-glycosylation–GalNAc-T isoforms direct distinct subsets of the *O*-glycoproteome. *EMBO Rep.* **16,** 1713–1722 CrossRef Medline
43. Schjoldager, K. T., Vester-Christensen, M. B., Bennett, E. P., Levery, S. B., Schwientek, T., Yin, W., Blixt, O., and Clausen, H. (2010) *O*-Glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. *J. Biol. Chem.* **285,** 36293–36303 CrossRef Medline
44. Fortelny, N., Yang, S., Pavlidis, P., Lange, P. F., and Overall, C. M. (2015) Proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res.* **43,** D290–D297 CrossRef Medline
45. Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **51,** 299–321 Medline
46. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., and Clausen, H. (2011) Mining the *O*-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods* **8,** 977–982 CrossRef Medline
47. Lange, P. F., Huesgen, P. F., Nguyen, K., and Overall, C. M. (2014) Annotating N termini for the human proteome project: N termini and Nα-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. *J. Proteome Res.* **13,** 2028–2044 CrossRef Medline
48. Prudova, A., auf dem Keller, U., Butler, G. S., and Overall, C. M. (2010) Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degra-

domes by iTRAQ-TAILS quantitative proteomics. *Mol. Cell Proteomics* **9,** 894–911 CrossRef Medline

49. O'Donoghue, A. J., Jin, Y., Knudsen, G. M., Perera, N. C., Jenne, D. E., Murphy, J. E., Craik, C. S., and Hermiston, T. W. (2013) Global substrate profiling of proteases in human neutrophil extracellular traps reveals consensus motif predominantly contributed by elastase. *PLoS One* **8,** e75141 CrossRef Medline
50. Wang, L., Yao, M., Pan, L. H., Qian, Q., and Yao, D. F. (2015) Glypican-3 is a biomarker and a therapeutic target of hepatocellular carcinoma. *Hepatobiliary Pancreat. Dis. Int.* **14,** 361–366 CrossRef Medline
51. Petrovan, R. J., Govers-Riemslag, J. W., Nowak, G., Hemker, H. C., Tans, G., and Rosing, J. (1998) Autocatalytic peptide bond cleavages in prothrombin and meizothrombin. *Biochemistry* **37,** 1185–1191 CrossRef Medline
52. Nesheim, M. E., Abbott, T., Jenny, R., and Mann, K. G. (1988) Evidence that the thrombin-catalyzed feedback cleavage of fragment 1.2 at $Arg^{154}$-$Ser^{155}$ promotes the release of thrombin from the catalytic surface during the activation of bovine prothrombin. *J. Biol. Chem.* **263,** 1037–1044 Medline
53. Martínez-Oliván, J., Arias-Moreno, X., Velazquez-Campoy, A., Millet, O., and Sancho, J. (2014) LDL receptor/lipoprotein recognition: endosomal weakening of ApoB and ApoE binding to the convex face of the LR5 repeat. *FEBS J.* **281,** 1534–1546 CrossRef Medline
54. Fortelny, N., Cox, J. H., Kappelhoff, R., Starr, A. E., Lange, P. F., Pavlidis, P., and Overall, C. M. (2014) Network analyses reveal pervasive functional regulation between proteases in the human protease web. *PLoS Biol.* **12,** e1001869 CrossRef Medline
55. Shirakabe, K., Omura, T., Shibagaki, Y., Mihara, E., Homma, K., Kato, Y., Yoshimura, A., Murakami, Y., Takagi, J., Hattori, S., and Ogawa, Y. (2017) Mechanistic insights into ectodomain shedding: susceptibility of CADM1 adhesion molecule is determined by alternative splicing and *O*-glycosylation. *Sci. Rep.* **7,** 46174 CrossRef Medline
56. Duguay, S. J., Jin, Y., Stein, J., Duguay, A. N., Gardner, P., and Steiner, D. F. (1998) Post-translational processing of the insulin-like growth factor-2 precursor: analysis of *O*-glycosylation and endoproteolysis. *J. Biol. Chem.* **273,** 18443–18451 CrossRef Medline
57. Kingsley, D. M., Kozarsky, K. F., Segal, M., and Krieger, M. (1986) Three types of low density lipoprotein receptor-deficient mutant have pleiotropic defects in the synthesis of *N*-linked, *O*-linked, and lipid-linked carbohydrate chains. *J. Cell Biol.* **102,** 1576–1585 CrossRef Medline
58. Minond, D., Cudic, M., Bionda, N., Giulianotti, M., Maida, L., Houghten, R. A., and Fields, G. B. (2012) Discovery of novel inhibitors of a disintegrin and metalloprotease 17 (ADAM17) using glycosylated and non-glycosylated substrates. *J. Biol. Chem.* **287,** 36473–36487 CrossRef Medline
59. Boskovski, M. T., Yuan, S., Pedersen, N. B., Goth, C. K., Makova, S., Clausen, H., Brueckner, M., and Khokha, M. K. (2013) The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature* **504,** 456–459 CrossRef Medline
60. Overall, C. M. (2002) Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol. Biotechnol.* **22,** 51–86 CrossRef Medline
61. Stavenhagen, K., Hinneburg, H., Thaysen-Andersen, M., Hartmann, L., Varón Silva, D., Fuchser, J., Kaspar, S., Rapp, E., Seeberger, P. H., and Kolarich, D. (2013) Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J. Mass Spectrom.* **48,** 627–639 Medline
62. Anthony, R. M., Nimmerjahn, F., Ashline, D. J., Reinhold, V. N., Paulson, J. C., and Ravetch, J. V. (2008) Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* **320,** 373–376 CrossRef Medline
63. Wasser, C. R., and Herz, J. (2017) Reelin: neurodevelopmental architect and homeostatic regulator of excitatory synapses. *J. Biol. Chem.* **292,** 1330–1338 CrossRef Medline
64. Tseng, W. L., Chen, T. H., Huang, C. C., Huang, Y. H., Yeh, C. F., Tsai, H. J., Lee, H. Y., Kao, C. Y., Lin, S. W., Liao, H. R., Cheng, J. C., and Tseng, C. P. (2014) Impaired thrombin generation in Reelin-deficient mice: a potential role of plasma Reelin in hemostasis. *J. Thromb. Haemost.* **12,** 2054–2064 CrossRef Medline
65. Ding, Y., Huang, L., Xian, X., Yuhanna, I. S., Wasser, C. R., Frotscher, M., Mineo, C., Shaul, P. W., and Herz, J. (2016) Loss of Reelin protects against atherosclerosis by reducing leukocyte-endothelial cell adhesion and lesion macrophage accumulation. *Sci. Signal.* **9,** ra29 CrossRef Medline
66. Kohno, T., Honda, T., Kubo, K., Nakano, Y., Tsuchiya, A., Murakami, T., Banno, H., Nakajima, K., and Hattori, M. (2015) Importance of Reelin C-terminal region in the development and maintenance of the postnatal cerebral cortex and its regulation by specific proteolysis. *J. Neurosci.* **35,** 4776–4787 CrossRef Medline
67. Ha, S., Tripathi, P. P., Mihalas, A. B., Hevner, R. F., and Beier, D. R. (2017) C-Terminal region truncation of RELN disrupts an interaction with VLDLR, causing abnormal development of the cerebral cortex and hippocampus. *J. Neurosci.* **37,** 960–971 CrossRef Medline
68. Joshi, H. J., Jørgensen, A., Schjoldager, K. T., Halim, A., Dworkin, L. A., Steentoft, C., Wandall, H. H., Clausen, H., and Vakhrushev, S. Y. (2018) GlycoDomainViewer: a bioinformatics tool for contextual exploration of glycoproteomes. *Glycobiology* **28,** 131–136 CrossRef Medline
69. Lange, P. F., Huesgen, P. F., and Overall, C. M. (2012) TopFIND 2.0: linking protein termini with proteolytic processing and modifications altering protein function. *Nucleic Acids Res.* **40,** D351–D361 CrossRef Medline
70. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11,** 1453–1459 CrossRef Medline
71. Maddelein, D., Colaert, N., Buchanan, I., Hulstaert, N., Gevaert, K., and Martens, L. (2015) The iceLogo web server and SOAP service for determining protein consensus sequences. *Nucleic Acids Res.* **43,** W543–W546 CrossRef Medline
72. Bagdonaite, I., Nordén, R., Joshi, H. J., Dabelsteen, S., Nyström, K., Vakhrushev, S. Y., Olofsson, S., and Wandall, H. H. (2015) A strategy for *O*-glycoproteomics of enveloped viruses: the *O*-glycoproteome of herpes simplex virus type 1. *PLoS Pathog.* **11,** e1004784 CrossRef Medline
73. Bagdonaite, I., Nordén, R., Joshi, H. J., King, S. L., Vakhrushev, S. Y., Olofsson, S., and Wandall, H. H. (2016) Global mapping of *O*-glycosylation of varicella zoster virus, human cytomegalovirus, and Epstein-Barr virus. *J. Biol. Chem.* **291,** 12014–12028 CrossRef Medline
74. Spitzer, M., Wildenhain, J., Rappsilber, J., and Tyers, M. (2014) BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **11,** 121–122 CrossRef Medline
75. Klein, T., Fung, S. Y., Renner, F., Blank, M. A., Dufour, A., Kang, S., Bolger-Munro, M., Scurll, J. M., Priatel, J. J., Schweigler, P., Melkko, S., Gold, M. R., Viner, R. I., Régnier, C. H., Turvey, S. E., and Overall, C. M. (2015) The paracaspase MALT1 cleaves HOIL1 reducing linear ubiquitination by LUBAC to dampen lymphocyte NF-$\kappa$B signalling. *Nat. Commun.* **6,** 8777 CrossRef Medline
76. Lau, H. T., Suh, H. W., Golkowski, M., and Ong, S. E. (2014) Comparing SILAC- and stable isotope dimethyl-labeling approaches for quantitative proteomics. *J. Proteome Res.* **13,** 4164–4174 CrossRef Medline
77. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., *et al.* (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41,** D1063–D1069 Medline
78. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 CrossRef Medline
79. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 CrossRef Medline
80. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6,** e21800 CrossRef Medline
81. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–D452 CrossRef Medline